HE 20.7102: Om 8

NOV 2 - 1983

103

Dealing With Outlying Observations

556

ORIGINAL

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
CENTERS FOR DISEASE CONTROL
NATIONAL INSTITUTE FOR OCCUPATIONAL SAFETY AND HEALTH

NIOSH

DIVISION OF TRAINING AND MANPOWER DEVELOPMENT

BEST COPY AVAILABLE

DEALING WITH OUTLYING OBSERVATIONS

–556–

Division of Training and Manpower Development

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control
National Institute for Occupational Safety and Health
Cincinnati, Ohio

August 1983

I

BEST COPY AVAILABLE

DISCLAIMER

Mention of company names or products does not constitute endorsement by the National Institute for Occupational Safety and Health.

BEST COPY AVAILABLE

FOREWORD

This text contains course material for ten lessons of the Dealing with Outlying Observations Course Package. The material covered in this course provides a number of techniques for the treatment of outliers, depending on their number and place in the array of observed data, as well as problems and exercises offering practice in the use of such techniques.

BEST COPY AVAILABLE

BLANK PAGE

CONTENTS

Foreword iii

LESSON 1
REGISTRATION AND ORIENTATION 1

LESSON 2
INTRODUCTION TO OUTLYING OBSERVATIONS – OVERVIEW 5

LESSON 3
APPLICATIONS AND EXERCISES – HYPOTHESIS TESTING REVIEW 19

LESSON 4
SINGLE OUTLIERS 23

LESSON 5
EXAMPLES AND EXERCISES – SINGLE OUTLIERS 39

LESSON 6
MULTIPLE OUTLIERS 45

LESSON 7
EXAMPLES AND EXERCISES – MULTIPLE OUTLIERS 69

LESSON 8
CONTINUOUS DATA BEARING OUTLIERS 73

LESSON 9
EXAMPLES AND EXERCISES – CONTINUOUS DATA OUTLIERS 85

LESSON 10
REVIEW AND EVALUATION 91

BEST COPY AVAILABLE

Lesson 1

REGISTRATION AND ORIENTATION

TIME: First Period, First Day (30 minutes)

OBJECTIVES: Enable the trainee to fill out the registration form correctly and to use the course materials for maximum learning efficiency.

EXHIBITS: None

MATERIALS: Student Manual

Registration Form

Course Information Packet

METHOD: Explanation/Lecture

BEST COPY AVAILABLE

ABSTRACT

The following is an outline of the contents of the Registration and Orientation lesson:

- Introduction
- Course Schedule
- Lessons
  - Format
  - Practical Exercises
  - Audio-Visual Aids
- Study Tips

BEST COPY AVAILABLE

Lesson 1

REGISTRATION AND ORIENTATION

INTRODUCTION

Welcome to the Dealing with Outlying Observations Course. This course is offered by the NIOSH Division of Training and Manpower Development to help fulfill the responsibilities of the U.S. Department of Health and Human Services under the Occupational Safety and Health Act of 1970. The course was developed to supplement the statistical techniques discussed in Courses 587 and 553, "Industrial Hygiene Laboratory Quality Control" and "Industrial Hygiene Sampling, Decision-Making, Monitoring and Record-keeping: Sampling Strategies." The course is part of a NIOSH program and trains industrial hygienists and others who are concerned with the quality of laboratory results.

The purpose of this course is to teach laboratory scientists, analysts and technicians how to deal with and evaluate outlying observations in a statistically valid manner.

This lesson gives you a chance to become acquainted with various aspects of our course. The instructor will give you a briefing on how the course will be conducted. Now is the time to ask any questions you may have about course materials, facilities, etc.

COURSE SCHEDULE

This course is designed for completion in two full days of classwork. The course consists of 10 lessons of varying length. The schedule allows a 10-minute break between each session and a 1 hour lunch break. The schedule is not meant to be followed rigidly and may be modified to satisfy practical work or discussion needs.

LESSONS

The lessons feature either a lecture or a workshop in which trainees have an opportunity to solve problems involving key points previously learned. The final class session is devoted to a course evaluation, and an informal question and answer period.

FORMAT

Each of the 10 lessons in the course begins with an "overview" page that tells you what to expect in that part of the instruction. The overview of each lesson gives you important information about:

- Lesson name and number
- Lesson objectives
- Exhibits or visuals found in the lesson
- Special words or "key terms" of importance in the lesson
- Materials needed
- Methods of instruction

The overview includes an outline of the contents of the lesson. The rest of the lesson usually contains explanation/lecture notes,

BEST COPY AVAILABLE

exercises, or other related instructional material. Most lessons close with a list of the references that had been noted in the text or that might provide further information to an interested student on a particular subject.

Workshops

Workshop lessons contain a number of exercises, followed by the solutions to the exercises. You are allowed to look back through the text, as needed, to answer these exercise questions. The problem sets are also followed by their solutions so you can check your work and spot where you are having difficulties.

Audio-Visual Aids

The instructor will show transparencies in many of the lessons. Copies of these transparencies are found in the text material and are referred to as Exhibits. All Exhibits are titled and numbered for easy reference.

BEST COPY AVAILABLE

LESSON 2

INTRODUCTION TO OUTLYING OBSERVATIONS – OVERVIEW

TIME: Second Period, First Day (50 minutes)

OBJECTIVES: Enable the trainee to:

1. Understand the historical concern about the rational treatment of outliers
2. Understand the relationship of hypothesis testing to the treatment of outliers

EXHIBITS: Figure 2.1. Example Data Sets

Figure 2.2. Population Distributions

MATERIALS: Student Manual

Pencils and Paper

Chalk Board, Chalk and Eraser

Overhead Projector and Screen

METHOD: Explanation/Lecture

BEST COPY AVAILABLE

ABSTRACT

The following is an outline of the contents of the Hypothesis Testing Review lesson:

- Introduction
- Historical Background
- Overview of Outlier Problems
- Base Population Characteristics
- The Choice of Outlier Rejection Tests
- Hypothesis Testing Review
- Summary
- References

BEST COPY AVAILABLE

Lesson 2

INTRODUCTION TO OUTLYING OBSERVATIONS

INTRODUCTION

Dealing with outlying observations or "outliers" is one of the most frustrating and perplexing problems facing the industrial hygiene program members. The writings of some "strict" statisticians would lead us to a position of "never rejecting an observation" unless a physical cause can be documented. In that case, we don't need statistics anyway. The other end of the spectrum is the temptation to reject all "bad looking" or "deviant" observations summarily. Between these extremes we must operate to assure a clear conscience and a statistically sound program. After looking at some historical background, we will discuss the philosophical problems, the scope of possible potential problems and the use of hypothesis testing.

A problem arising from the many sources of references for this and subsequent lessons is the use of a variety of statistical notations by authorities cited. To attempt to minimize confusion NIOSH has selected the following as "standard" for this course:

- σ — Population Standard Deviation
- s — Sample Standard Deviation
- $\bar{X}$ — Sample Mean
- μ — Population Mean

I. HISTORICAL BACKGROUND

The treatment of outlier observations is a problem as old as data itself. Procedures to assure reproducible testing and thereby avoid the likelihood of outliers go back at least as far as the middle ages. Galileo's work on the acceleration of gravity in the 1600's and Milliken's work of about 1900 on determining the surface tension T of a liquid are classical recorded examples. (Eisenhart)[4]

Walter A. Shewhart in 1939 (Ibid)[4] was concerned with the "reproducibility of the operation." "Hence, it seems to me that the characteristic of the numerical aspects of an operation that is of greatest practical interest is its *reproducibility within tolerance limits throughout the "infinite sequence."* The limit to which we may go in this direction is to attain a state of statistical control. The attempt to attain a certain kind of consistency within the first n observed values is merely a means of attaining reproducibility within limits throughout the whole of the sequence."

The entire purpose of Shewhart's message is to reinforce the basic tenet of probability theory. Only if those "n" measurements are a random sample from the population or universe of conceivable measurements can they be useful in describing the population or universe. The mathematical form of the probability distribution is not limited, but there must be, indeed, a form. Deming in 1950 reiterated the reality of life when he said (Ibid) "In applying statistical theory, the main consideration is not what the shape of the universe is, but whether there is any universe at all. No universe can be assumed, nor . . . statistical theory . . . applied unless the observations show statistical control. In this state of statistical control the samples when cumulated over a suitable interval of time give a distribution of a particular shape, and this shape is reproduced hour after hour, day

BEST COPY AVAILABLE

after day, so long as the process remains in statistical control—i.e., exhibits the properties of randomness. In a state of control, n observations may be regarded as a sample from the universe of whatever shape it is. A big enough sample, or enough small samples, enables the statistician to make meaningful and useful predictions about future samples. This is as much as statistical theory can do.

" . . . Very often the experimenter, instead of rushing in to apply (statistical methods) should be more concerned about attaining statistical control and asking himself whether any predictions at all (the only purpose of his experiment), by statistical theory or otherwise, can be made."

One cannot, without attendant risk, assume that a set of observations are indeed a random sample from a population. Shewhart in 1940 (ibid) at a University of Pennsylvania Conference gave practical criteria. "One should consider a measurement process to be — i.e., should "decide to act for the present as if "the process were — *in a state of (simple) statistical control*, only if the measurements in hand show no evidence of lack of statistical control when analyzed for randomness *in the order in which they were* taken by the control chart techniques for averages and standard deviations that he had found so valuable in industrial process control and by certain additional tests for randomness based on "runs above and below average" and "runs up and down."

A few comments on the "average" or mean are also in order. Thomas Simpson (Ibid) in a 1755 letter to the President of the Royal Society of London "On the Advantage of Taking the Mean of a Number of Observations, in Practical Astronomy," was the first to consider repeated measurements of a simple quantity by given measurements of a single quantity by a given measurement process as observed values of independent random variables having the same probability distribution. Simpson said "Upon the whole of which it appears, that the taking of the mean of a number of observations, greatly diminishes the chances for all the smaller errors, and cuts off almost all possibility of any great ones: which last consideration alone, seems sufficient to recommend the use of the method, not only to astronomers, but to all others concerned in making of experiments of any kind (to which the above reasoning is equally applicable). And the more observations or experiments there are made, the less will the conclusion be liable to err, provided they admit of being repeated under the same circumstances.

Simpson brought the subject of errors of measurement into the arena of mathematical analysis as being distributed in accordance with a probability distribution that was an intrinsic property of the measurement process itself. Thus, the theory of measurement is based on the theory of probability.

Student (William Gosset, 1867—1937) is recognized as the Father of the "Theory of small samples." (ibid) He stressed the importance of randomness in measurement and experimentation in 1908[19] with these remarks, "Any experiment may be regarded as forming an individual of a population of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population. Now any series of experiments is only of value in-so-far as it enables us to form a judgement as to the statistical constants of the population to which the experiments belong."

The attainment of an ideal strict statistical control may be difficult if not impossible to achieve. The economic feasibility may negate practicability even if technically feasible. However, the observer or data analyst assumes greater and greater risks in accepting the assumptions that underlie the tests for outliers, the further the deviation from statistical control.

R. B. Murphy of Bell Labs in 1961 phrased the dilemma we face when he said . . . "an effort to follow a test method ought not to be known as a measurement process unless it is capable of statistical control. Capability of control

BEST COPY AVAILABLE

means that either the measurements are the product of an identifiable statistical universe or an orderly array of such universes or, if not, the physical causes preventing such identification may themselves be identified and, if desired, isolated and suppressed. Incapability of control implies that the results of measurement are not to be trusted as indications of the physical property at hand; in short, we are not in any verifiable sense measuring anything."

II. OVERVIEW OF OUTLIER PROBLEMS

The detection of outliers process is used for one of three purposes (David)[3]: a) screen routine data sets prior to statistical analysis; b) determine if the process being monitored has changed significantly in mean or variance; or c) detect extreme values which may result from improvements due to process changes such as engineering controls.

Most of the traditional outlier tests have been devised for looking at a single population. Slippage tests are developed to test the equality of several populations with the alternative hypothesis that one of the populations is different or has "slipped." When each population is defined by only a single observation, the classical detection of outlier techniques apply to this limiting case.

Purposes b and c above can be related to process quality control and investigational situations. Fleming's discovery of penicillin is such a case. The screening of a set(s) of data prior to statistical analysis is the area of usual concern. Figure 2.1 illustrates three common situations.

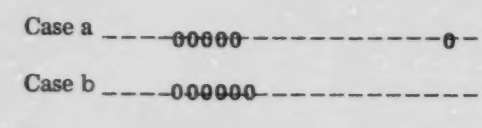


Figure 2.1. Example Data Sets

Case a and Case b seem to be clearcut decisions intuitively. Case c, of course, bothers us the most. Before we worry about "how" to deal with this case, we should consider the "why." What is the cause of the suspected outlier? What is the nature of the population(s) that we are dealing with?

Figure 2.2 shows two potential populations.

The situations described in Figure 2.2 are both routinely encountered by the industrial hygiene program. Case (a) may well arise from equipment or other malfunctioning aspects of either the sampling or analytical aspects of the measurement process. Case (b) represents a reality of the workplace environment in many cases. Much of the work of NIOSH in its *Occupational Exposure Sampling Strategy Manual* (NIOSH)[8] is based upon this assumption (Leidel).[8]

It is well to again recall that the principle and preferred technique is to determine the physical or chemical cause of the outlier observation (ASTM)[1] and thereby make the rejection decision.

Regardless of the protocol used for rejection of outliers it is important to acknowledge their presence in the original data and describe the rejection criteria. "It is much better to state their values and to do alternative analysis using all or some of them" (Krushal).[6] Another caution relates to the ultimate use of the data being generated. If, after testing a workplace atmosphere, one of the five measurements is obviously out of line, we would reject it. Such rejection, especially if the cause, such as an equipment malfunction, is discovered can be justified. However, if we are interested in determining the effectiveness of the monitoring procedure, the rejected observation and its cause are a vital part of the process reality.

The warning issued by Pearson and Chandrasehar in 1936 and quoted by K. R. Nair in 1948 (Nair)[17] bears repeating here. "To base the choice of the test of a statistical hypothesis

BEST COPY AVAILABLE

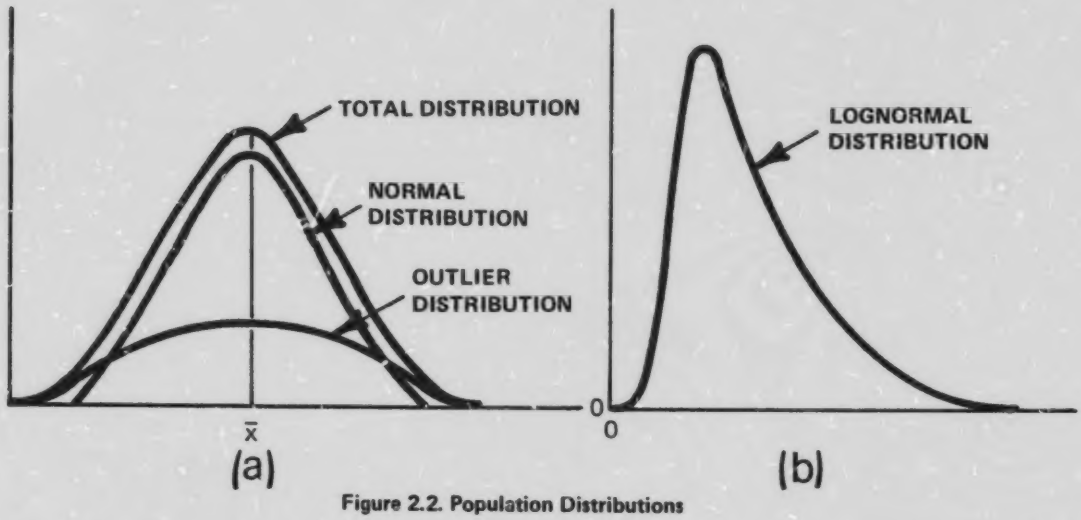


Figure 2.2. Population Distributions

upon an inspection of the observations is a dangerous practice; a study of the configuration of a sample is almost certain to reveal some feature, or features which are exceptional if the hypothesis is true. By choosing the feature most unfavorable to the hypothesis out of a very large number of features examined, it will usually be possible to find some reason for rejecting the hypothesis."

III. BASE POPULATION CHARACTERISTICS

In the ideal world the variability associated with our measurement processes would be described by the Normal or Gaussian distribution. In industrial hygiene very many measurement situations have been better described (at least empirically) by the log-normal distribution. It is not necessary to know the causes of this situation, even though that would assist us in understanding a data set and dealing with the outliers. One can postulate that the lower end is truncated by zero or at least by the lower limit of detection of the measurement method.

Other distributions for the population might be observed as well. The practical problem in outlier rejection is that the validity or efficiency of the test is different for differently shaped populations. The difference may be significant.

There are at least three approaches that can be used in dealing with this problem. The simplest involves the Central Limit Theorem which has been described by the following: "If a population has a finite variance σ^2 and mean μ, then the distribution of the sample mean (of n independent measurements) approaches the normal distribution with variance σ^2 /n and mean m as sample size n increases. This remarkable and powerful theorem is indeed tailored for measurement processes" (Ku).[7] The central limit theorem thereby provides an effective normalization procedure for data even when the population distribution is markedly non-normal.

The second approach is the use of normalizing transformations. When we know, from previous experience or on a theoretical basis, that a transformation will approximately normalize the data set, it may be more efficient and convenient to use normal-based outlier tests than distribution free tests (Natrella, Chapter 20).[11]

BEST COPY AVAILABLE

The third approach is the use of distribution free tests for outliers. One such approach is the use of ranks (Wonnacott)[13] or (Conover).[2]

IV. THE CHOICE OF THE OUTLIER REJECTION TESTS

The choice of which test to apply to examine a set of data for rejection of a single outlier is based upon several considerations. The first consideration is whether the base population is known or assumed to be normal. The following listings of tests are based upon that assumption. The second consideration is how much is known about μ and σ. The third consideration is whether we are interested in shifts in a specified direction (one-sided) or in either direction (two-sided). Since in the majority of industrial hygiene applications, the suspect outlier undergoing scrutiny will be on the right or upper side, that formulation is given for the one-sided tests.

The following listings of tests are based upon David, pp. 172–3.[3]

The choice between the available tests appropriate to the sets of conditions listed above will be discussed in later sections of the course.

V. OUTLIER TESTS WHEN μ AND σ ARE KNOWN

The case where both μ and σ are known is the basic theoretical situation. Industrial Hygiene programs will encounter the situation rarely if at all in routine practice.

The first and most basic test is based upon the use of the Cumulative Normal Distribution — values of (Table 2.1) such as given in NBS Handbook 91 page T-3 (Natrella)[10] and other standard sources. This Table is reproduced as Table 2.1. The test follows Natrella Ch. 17.[10]

The second test involves the use of the Chi-Square (χ^2) Test (Natrella Ch. 9).[9]. The specific comparison would be that of comparison to a theoretical standard (the cumulative normal distribution). Several categories are used to describe the population characteristics. This test is of course usually used for a broader test of normality and the use to detect outliers is just one application. If one uses a minimum number of cells such as three (3) for a one-sided test the cell limits could be set as from -8 to 0.5 to .95; .95 to $+8$. Any other set of rational groups could be set up. The use of a two-sided test would require the cells to include one at each tail of the distribution.

VI. SUMMARY

In the outline we have reviewed the philosophy of outlier rejection as well as the practical limitations. Outlier rejection techniques have been derived from both a historical and theoretical basis. A survey of selected techniques illustrated the limitations based upon the assumption of a normal population and differing levels of information about the population estimates of μ and σ. Several techniques for outliers illustrative of the most knowledge (both μ and σ known) apply the principles covered earlier.

BEST COPY AVAILABLE

One-Sided	*Two-Sided*
When both μ and σ are known	
$A_1 = \frac{(X_{(n)} - \mu)}{\sigma}$	$A_2 = \max \frac{\lvert X_i - \mu \rvert}{\sigma}$
Pearson & Hartley 1966[12] Table 24	$A_3 = \chi^2_n = \frac{\Sigma (Xi - \mu)^2}{\sigma^2}$
When only σ is known	
$B_1 = \frac{(X_{(n)} - \bar{X})}{\sigma}$	$B_2 = \max \frac{\lvert X_i - \bar{X} \rvert}{\sigma}$
Grubbs (1950)[5]	
$B_3 = \frac{(X_{(n)} - X_{(n-1)})}{\sigma}$	$B_4 = \frac{\text{Range}}{\sigma}$
Irwin, 1925[16]	Pearson & Hartley 1966 Table 22
	$B_5 = \chi^2_{n-1} = \frac{\Sigma (X_i - \bar{X})^2}{\sigma^2}$
Both μ and σ unknown but have an Independent Estimate of σ^2	
$C_1 = \frac{(X_n - \bar{X})}{S_v}$	$C_2 = \max \frac{\lvert X_i - \bar{X} \rvert}{S_v}$
Where S_v^2 is an independent mean square estimate of σ^2 based on V Degrees of Freedom.	
Pearson & Hartley Table 26	Halperin et.al. 1955[13]
	$C_3 = \frac{\text{Range}}{S_v}$
	Pearson & Hartley Table 29
$C_4 = \frac{(X_{(n)} - \bar{X})}{[\Sigma (X_i - \bar{X})^2 + VS_v^2]^{1/2}}$	$C_5 = \frac{\max \lvert X_i - \bar{X} \rvert}{[\Sigma (X_i - X)^2 + VS_v^2]^{1/2}}$
Pearson & Hartley Table 26a	Pearson & Hartley Table 26b

BEST COPY AVAILABLE

One-Sided

Two-Sided

Both μ and σ Unknown

$$D_1 = \frac{(X_{(n)} - \bar{X})}{\frac{[\Sigma(X_i - \bar{X})^2]^{1/2}}{n}}$$

Grubbs 1950

$$D_2 = \frac{\text{Max } |X_i - \bar{X}|}{\frac{[\Sigma(X_i - \bar{X})^2]^{1/2}}{n-1}}$$

Pearson & Hartley
Table 26b

$$D_3 = \frac{\text{Range}}{\frac{[\Sigma(X_i - \bar{X})^2]^{1/2}}{n-1}}$$

Pearson & Hartley
Table 26b

$$D_4 = \frac{n^{1/2}\Sigma(X_i - \bar{X})^3}{[\Sigma(X_i - \bar{X})^2]^{3/2}}$$

Pearson & Hartley
Table 34b

$$D_5 = \frac{n\Sigma(X_i - \bar{X})^4}{[\Sigma(X_i - \bar{X})^2]^2}$$

Pearson & Hartley
Table 34c

$$D_6 = \frac{\sum_{i=1}^{n-2}\left[X_{(i)} - \left(\frac{\sum_{i=1}^{n-2}}{n-2}\right)X_{(i)}\right]^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Grubbs, 1950

D_7 = Dixon's r statistics, Dixon 1951

BEST COPY AVAILABLE

Procedure

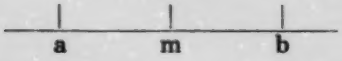
1. Choose α, the probability of falsely rejecting an observation when all observations do belong to the same group.

2. Compute $\alpha^1 = 1 - (1 - \alpha)^{1/n}$
   Use logarithms or a table of fractional powers.

3. Look up $Z_{1-\alpha^{1/2}}$ in Table 2.1. It may be necessary to interpolate in Table 2.1. Use of probability paper is recommended.

4. Compute: $a = m - \sigma Z_{1-\alpha^{1/2}}$
   $b = m + \sigma Z_{1-\alpha^{1/2}}$

5. Any observation not in the interval a-b as shown below



is rejected

WHERE m = the population mean

Example

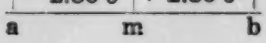
1. Let $\alpha = 0.10$ or other predetermined value.

2. If n = 20
   $\alpha^1 = 1 - (1 - .10)^{1/20}$
   $= 1 - (.90)^{1/20}$
   $= 1 - .9947$
   $= .0053$

3. $1 - \alpha^{1/2} = 1 - (0.0053/2)$
   $= .9974$
   $Z_{.9974} = 2.80$

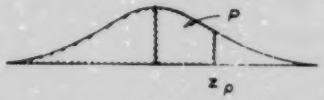
4. $a = m - 2.80\,\sigma$
   $b = m + 2.80\,\sigma$

5. Reject an observation not in the interval
   $m - 2.80\,\sigma$ to
   $m + 2.80\,\sigma$ as shown below



BEST COPY AVAILABLE

TABLE 2.1. CUMULATIVE NORMAL DISTRIBUTION — VALUES OF z_p



Values of z_p corresponding to P for the normal curve.

z is the standard normal variable

P	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.00	—	−2.33	−2.05	−1.88	−1.75	−1.64	−1.55	1.48	1.41	1.34
.10	−1.28	−1.23	−1.18	−1.13	−1.08	−1.04	−0.99	0.95	0.92	0.88
.20	−0.84	−0.81	−0.77	−0.74	−0.71	−0.67	0.64	0.61	0.58	0.55
.30	−0.52	−0.50	−0.47	−0.44	−0.41	−0.39	−0.36	0.33	0.31	0.28
.40	−0.25	−0.23	−0.20	−0.18	−0.15	−0.13	−0.10	0.08	0.05	0.03
.50	0.00	0.03	0.05	0.08	0.10	0.13	0.15	0.18	0.20	0.23
.60	0.25	0.28	0.31	0.33	0.36	0.39	0.41	0.44	0.47	0.50
.70	0.52	0.55	0.58	0.61	0.64	0.67	0.71	0.74	0.77	0.81
.80	0.84	0.88	0.92	0.95	0.99	1.04	1.08	1.13	1.18	1.23
.90	1.28	1.34	1.41	1.48	1.55	1.64	1.75	1.88	2.05	2.33

Special Values

P	.001	.005	.010	.025	.050	.100
z_p	−3.090	−2.576	-2.326	−1.960	−1.645	−1.282

P	.999	.995	.990	.975	.950	.900
z_p	3.090	2.576	2.326	1.960	1.645	1.282

BEST COPY AVAILABLE

REFERENCES

1. Dealing with Outlying Observations — E-178-68, 1972 Annual Book of ASTM *Standards*, American Society for Testing and Materials, Philadelphia, pp. 378—93, 1972.

2. Conover, W. J., *Practical Nonparametric Statistics*, John Wiley & Sons, New York, N.Y., pp. 342—49. 1971.

3. David, H. A., *Order Statistics*, John Wiley & Sons, Inc., New York, N.Y. 1970.

4. Eisenhart, Churchill, Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems, *Precision Measurement and Calibration-Statistical Concepts and Procedures*, Harry H. Ku, Editor, NBS Special Publication 300— Volume 1. Washington, D.C. 1969.

5. Grubbs, F. E., Sample Criteria for Testing Outlying Observations, Ann. Math. Statist. 21, 27—58. 1950.

6. Kruskal, William H., Some Remarks on Wild Observations, Technometrics, Volume 2, 1960.

7. Ku, Harry H., Statistical Concepts in Metrology, *Precision Measurement and Calibration — Statistical Concepts and Procedures*, Harry H. Ku, Editor, NBS Special Publication 300 — Volume 1, Washington, D.C., 1969.

8. Leidel, N. A., Busch, K. A., and Lynch, J. R., *Occupational Exposure Sampling Strategy Manual*, USDHEW, CDC, NIOSH, DHEW (NIOSH) Publ. #77-173, Cincinnati, Ohio 1977.

9. Natrella, Mary, Comparing Materials or Products with Respect to Several Categories of Performance (Chi-Square Tests) Chap. 9, *Experimental Statistics*, NBS Handbook 91, Washington, D.C. 1963.

10. Natrella, Mary, The Treatment of Outliers, Chap. 17, *Experimental Statistics*, NBS Handbook 91, Washington, D.C. 1963.

11. Natrella, Mary, The Use of Transformations, Chap. 20, *Experimental Statistics*, NBS Handbook 91, Washington, D.C. 1963.

12. Pearson, E. S., and Hartley, H. O., *Biometrika Tables for Statisticians*, Volume 1, 3rd Ed., Cambridge University Press, 1966.

13. Wonnacott, T. H. and Wonnacott, R. J., *Introductory Statistics*, John Wiley and Sons, Inc., New York, N.Y., 1977.

BEST COPY AVAILABLE

BIBLIOGRAPHY

14. Deming, W. Edward, *Some Theory of Sampling*, John Wiley & Sons, New York, New York, 1950.

15. Halperin, M., Greenhouse, S. W., Cornfield, J., and Zalohor, J., Tables of Percentage Points for Studentized Maximum Absolute Deviate in Normal Samples, J. Amer. Statist. Ass. 50, 185—95. 1955.

16. Irwin, J. O., On a Criteria for the Rejection of Outlying Observations, Biometrika 17, 238—50, 1925.

17. Nair, K. R., *The Distribution of the Extreme Deviate from the Sample Mean and Its Studentized Forms.* Biometrika Volume 35, 1948, pp. 114—118.

18. Shewhart, Walter A., *Statistical Method from the Viewpoint of Quality Control*, The Graduate School, U.S. Department of Agriculture, Washington, D.C. 1939.

19. Shewhart, Walter A., Contributions of Statistics to the Science of Engineering, University of Pennsylvania Bicentennial Conference Volume on *Fluid Mechanics and Statistical Methods in Engineering*, pp. 97—124, University of Pennsylvania Press, Philadelphia, 1941.

20. Student, The Probable Error of a Mean, Biometrika, VI, No. 1, pp. 1—25, 1908.

BEST COPY AVAILABLE

BLANK PAGE

LESSON 3

APPLICATIONS AND EXERCISES — HYPOTHESIS TESTING REVIEW

TIME: Third Period, First Day (50 minutes)

OBJECTIVES: Enable the trainee to complete a set of calculations so that he can exercise judgement as to the validity of single aberant observations.

EXHIBITS:

Figure 3.1. Solution to Calculations

Figure 3.2. Critical Value Diagram

Figure 3.3. Critical V lue Diagram

Figure 3.4. Critical V Diagram

MATERIALS:

Student Manual

Pencils with Erasers

Calculators with Log Func ns

Overhead Projector with Screen

Transparencies 1 and 2

BEST COPY AVAILABLE

ABSTRACT

This lesson provides practice in using the concepts introduced in Lesson 2. The following is an outline of the Applications and Exercises lesson:

- Introduction
- Exercises
- Solutions to Exercises

BEST COPY AVAILABLE

Lesson 3

APPLICATIONS AND EXERCISES — HYPOTHESIS TESTING REVIEW

INTRODUCTION

A review of hypothesis testing was presented in Lesson 2. In this lesson you will have an opportunity to use these tests in the decision-making process in dealing with data containing single outliers.

Use a calculator to compute logarithms and antilogs.

1. Work by yourself, unless the instructor directs otherwise.
2. Read and follow the material in the text in this and the previous lesson.
3. Ask the instructor for help if you need it.

EXERCISES

Complete the following exercises according to the directions for each one:

1. A Tedlar bag with a known concentration of carbon disulfide was prepared in the industrial hygiene laboratory. Twenty measurements were made on the contents. From previous experience it is known that the population mean is 10 ppm and the population standard deviation is 2 ppm.

   The measurements were: 10.1, 9.9, 7.2, 13.6, 12.7, 10.8, 6.6, 8.1, 11.1, 8.6, 5.8, 10.9, 11.0, 5.6, 13.4, 8.2, 8.9, 6.6, 9.7, and 14.6.

   Calculate the critical values to determine whether any values should be rejected as outliers.

2. Ten measurements of the concentrations of carbon monoxide in a Tedlar bag were made. From considerable previous experience it is known that the population mean is 100 ppm and the population standard deviation is 20 ppm.

   The values measured were 87, 78, 97, 58, 92, 103, 100, 106, 123 and 93.

   The question is: Should the values of 58 and 123 be ignored?

3. A second known concentration of carbon monoxide was prepared. It is known that the population mean is 50 ppm and the population standard deviation is 10 ppm.

   The values measured were 44, 21, 50, 46, 50, 46, 49, 40, 66, and 43. Should the values of 21 and 66 be ignored?

BEST COPY AVAILABLE

BLANK PAGE

LESSON 4

SINGLE OUTLIERS

TIME: Fourth Period, First Day (50 minutes)

OBJECTIVES: Enable the trainee to select the most appropriate test statistic to evaluate the significance of a single outlier depending on how much information about the data is available.

EXHIBITS:

Table 4.1 Critical Values of B_1 When the Population Standard Deviation σ is Known (Nair)

Table 4.2 Critical Values of B_1 When the Population Standard Deviation σ is Known (Grubbs)

Table 4.3 Test Statistic $B_4 = \frac{\text{Range}}{\sigma}$

Table 4.4 Test Statistic $C_1 = \frac{(X_n - \bar{X})}{s_v}$ (Nair)

Table 4.5 Test Statistic $C_1 = \frac{(X_n - \bar{X})}{s_v}$ (Grubbs)

Table 4.6A Test Statistic $C_3 = \frac{X_n - X_1}{s_v}$

Table 4.6B Test Statistic $C_3 = \frac{X_n - X_1}{s_v}$

Table 4.7 Test Statistic $D_1 = \frac{X_n - \bar{X}}{s}$ or $\frac{\bar{X} - X_1}{s}$

Critical values for D_1
When Standard Deviation is Calculated from the Same Sample.

Table 4.8 Test Statistic $D_3 = \frac{\text{Range (w)}}{s}$ or $\frac{X_n - X_1}{s}$

Table 4.9 Dixon Ratio Test Statistics

KEY TERMS: Location error

Scalor error

BEST COPY AVAILABLE

MATERIALS: Student Manual

Pencils and Paper

Chalk Board and Chalk Erasers

Overhead Projector and Screen

Overhead Transparencies of Tables 4.1—4.9

ABSTRACT

The following is an outline of the contents of the Single Outliers lesson:

- Introduction
- Population Contamination
- Testing When Only σ is Known
- Testing When Both μ and σ are Unknown but There is an Independent
- Estimate of σ^2.
- Testing When Both μ and σ are Unknown
- Summary
- References

BEST COPY AVAILABLE

Lesson 4

SINGLE OUTLIERS

INTRODUCTION

The first discussion of outliers set the stage for the commonplace reality of dealing with a suspect observation in a set of data. This set of data may be a sampling from a process where the variability has been so well established over time that the population variability σ can be said to be known. A second case involves less knowledge about the population parameters. Our knowledge of the variability is not so certain but we do have available an independent estimate of σ, s_ν. The second case is common in industrial hygiene practice where we have reasonably well established the variability of the sampling and analytical process or the variability of the direct reading instrument in similar situations. The third case is common also. We have a set of data from a new plant, a new process, a new instrument, or a new analytic method without any historical data to assist in interpretation. Both μ and σ are unknown.

This discussion will deal with the cases where "A"; μ is unknown and σ known; "B"; where μ is unknown, σ is unknown but there is an independent estimate, s_ν of σ; "C"; where both μ and σ are unknown. Only single outliers will be discussed at this time. The outlier may occur at either end of the distribution, a two-sided test; or may occur only at the right of the sample (the most common situation) or at the left of the sample. These are one-sided tests.

Prior to discussing the several tests available for each of these situations it would be well to reflect upon the types of happenings that cause the outlier we are testing for, or what conditions have contaminated the measurement process to cause the suspected outlier.

EXPLANATION/LECTURE

I. POPULATION CONTAMINATION

We will start with the assumption that the data set is a sample from a single normal population. Dixon[3] (1953) has provided a good characterization of the problem. The normal population has a mean μ and variance σ^2 and is described as follows $N(\mu, \sigma^2)$. Suppose that the population distribution is at some point in time disturbed due to a calibration shift, for example. The mean, μ is shifted by an increment described in terms of standard deviations from the mean or $\lambda\sigma$ where λ is the number of σ units of displacement. The population is now described (after the shift) by $N(N + \lambda\sigma, \sigma^2)$. A measurement or observation from this new population is called a location error. One could also experience an increase in the variability of the measurement process. A measurement or error from the more variable population $N(\mu, \lambda^2 \sigma^2)$ introduced into the original population $N(\mu,\sigma^2)$ is called a scalor error by Dixon.[2] We have added this dimension to the discussion since we will need to evaluate the efficiency of tests for outliers with respect to how well they perform when dealing with populations contaminated by either a displacement or scalor error.

II. WHEN ONLY σ IS KNOWN

The test B_1 and its derivatives were attributed to Grubbs 1950[4] and should also be attributed

BEST COPY AVAILABLE

to Nair 1948.[8] The test B_1 takes the form:

$$B_1 = \frac{(X_n - \bar{X})}{\sigma}$$

The two-sided version of this test B_2 takes the form:

$$B_2 = \max \frac{|X_i - \bar{X}|}{\sigma}$$

Nair points out that the Irwin[7] 1925 test statistic B_3 which uses the ratio:

$$B_3 = \frac{(X_n - X_{n-1})}{\sigma}$$

is not intuitively as good as the original form of B_1 as proposed by McKay[2] (b) in 1935. Nair also states that the range test for outliers proposed by Student[10] in 1927

$$\frac{(X_n - X_1)}{\sigma} \text{ or the } B_4 \text{ form } \frac{W_n}{\sigma}$$

will be the better test statistic if both X_1 and X_n are outliers. This case will be considered in the next lecture. David[1] 1970 agrees with the conclusion of Nair.

The test statistic B_5 or the χ^2 (Chi-Square) test can also be a useful tool in outlier rejection.

Grubbs[5] in 1969 advocated his 1950 version of the B_j test statistic. Dixon in 1950[2] tabulated the classification of outlier rejection tests as presented earlier by David 1970.[1] Dixon tested the various test statistics for both location and scalor errors. Dixon concluded that test B_1 gives the best performance if a single location error is present. The tests B_3 and B_4 are slightly easier to calculate and their performance is nearly as good as for B_1. However, when more than one outlier may be expected in a single sample the performances of B_1 and B_4 increase while B_3 decreases. Dixon's testing program produced the conclusion that test B_1 yields maximum performance but that for ease of computation, almost equal performance for a single outlier, about equal performance for two outliers and for protection for outliers either above or below the mean, test statistic B_4 is recommended.

The tables to be used for test statistic B_1 are available from either Nair 1948[8] or Grubbs 1950.[4] The table for test statistic B_4 is provided by Pearson and Hartley[9] 1956 Table 22.

III. BOTH μ AND σ UNKNOWN BUT HAVE AN INDEPENDENT ESTIMATE OF σ^2

This test situation involves less certain knowledge about the population variability. The test statistics tend to parallel the "B" test statistics. Test C_1 replaces the σ of B_1 with an independent estimate of σ or s_v. Test C_2 parallels the test B_2. Test statistic C_3 or $C_3 = W_n/S_v$ follows B_4. Test statistics C_4 and C_5 represent pooled studentized versions of B_1.

The power of this series of tests is dependent upon the degrees of freedom involved in calculation of S_v. As d.f. increases, their efficiency approaches that of test statistic B_1.

Tables for a series of ratio tests statistics have been compiled by Pearson and Hartley[9] 1956.

IV. BOTH μ AND σ UNKNOWN

This situation is the case dealt with quite often by industrial hygienists or chemists. The only information at hand is from the set of observations themselves. The array of tests available parallels that of the previous cases, but in addition, Dixon[3] 1953 has provided a series of range and sub-range ratio tests that can quickly deal with suspected outlier observations.

The efficiency of these tests parallels the earlier cases. Test statistic D_6 was designed to

BEST COPY AVAILABLE

deal with two outliers (both to the right) and will be discussed in the succeeding lecture.

The Dixon series of tests are one-sided. Ferguson (1961)[1(b)] has also studied a two sided version of the r_{10} test or $r_{10}(2)$. David 1970[1] concludes that the D_1 test is best for one sided tests when the number of observations (n) increase. For small n, D_1, D_4 and r_{10} are equivalent. For the two-sided tests D_2, D_5 and $r_{10(2)}$ are about equal in performance. Where there are two suspected outliers in a single set of data, test statistic D_5 out performs D_2. The Dixon ratio tests are not designed to handle this situation at all. Grubbs (DARCOM)[6] confirms the value of the D_1 test over the Dixon ratio test for a single outlier. The series of Dixon ratio tests follow:

	If X_n is Suspect	If X_1 is Suspect
r_{10} = n=3 to 7	$\frac{X_n - X_{n-1}}{X_n - X_1}$	$\frac{X_2 - X_1}{X_n - X_1}$
r_{11} = n=8 to 10	$\frac{X_n - X_{n-1}}{X_n - X_2}$	$\frac{X_2 - X_1}{X_{n-1} - X_1}$
r_{21} = n=11-13	$\frac{X_n - X_{n-2}}{X_n - X_2}$	$\frac{X_3 - X_1}{X_{n-1} - X_1}$
r_{22} = n=14 to 25	$\frac{X_n - X_{n-2}}{X_n - X_3}$	$\frac{X_3 - X_1}{X_{n-2} - X_1}$

TABLE 4.1. TEST STATISTIC $B_1 = \frac{(X_n - \bar{X})}{\sigma}$

Critical Values of B_1 When the Population Standard Deviation σ is Known

Size of sample n	Lower percentage points						Upper percentage points					
	0·1	0·5	1·0	2·5	5·0	10·0	10·0	5·0	2·5	1·0	0·5	0·1
3	0·03	0·06	0·09	0·14	0·20	0·29	1·50	1·74	1·95	2·22	2·40	2·78
4	0·09	0·16	0·20	0·27	0·35	0·45	1·70	1·94	2·16	2·43	2·62	3·01
5	0·16	0·25	0·30	0·38	0·47	0·58	1·83	2·08	2·30	2·57	2·76	3·17
6	0·23	0·33	0·38	0·48	0·56	0·68	1·94	2·18	2·41	2·68	2·87	3·28
7	0·30	0·40	0·46	0·56	0·65	0·76	2·02	2·27	2·49	2·76	2·95	3·36
8	0·36	0·47	0·53	0·62	0·72	0·84	2·09	2·33	2·56	2·83	3·02	3·43
9	0·41	0·53	0·59	0·69	0·78	0·90	2·45	2·39	2·61	2·88	3·07	3·48

This table is taken from the paper of Nair 1948, page 140. By permission. Biometrika, Vol. 35, 1948.

BEST COPY AVAILABLE

TABLE 4.2. TEST STATISTIC $B_1 = \frac{(X_n - \bar{X})}{\sigma}$

Critical Values of $B_{1\,(1)} = \frac{(\bar{X} - X_1)}{\sigma}$ or $B_{1\,(n)} = \frac{(X_n - \bar{X})}{\sigma}$

When the Population Standard Deviation σ is Known

Number of Observations n	5% Significance Level	1% Significance Level	0.5% Significance Level
2	1.39	1.82	1.99
3	1.74	2.22	2.40
4	1.94	2.43	2.62
5	2.08	2.57	2.76
6	2.18	2.68	2.87
7	2.27	2.76	2.95
8	2.33	2.83	3.02
9	2.39	2.88	3.07
10	2.44	2.93	3.12
11	2.48	2.97	3.16
12	2.52	3.01	3.20
13	2.56	3.04	3.23
14	2.59	3.07	3.26
15	2.62	3.10	3.29
16	2.64	3.12	3.31
17	2.67	3.15	3.33
18	2.69	3.17	3.36
19	2.71	3.19	3.38
20	2.73	3.21	3.39
21	2.75	3.22	3.41
22	2.77	3.24	3.42
23	2.78	3.26	3.44
24	2.80	3.27	3.45
25	2.81	3.28	3.46

This table is taken from the paper of Grubbs 1969, page 19.
By permission.

BEST COPY AVAILABLE

TABLE 4.3. TEST STATISTIC $B_4 = \frac{\text{Range}}{\sigma}$

Percentage points of the distribution of the range

Size of sample n	Factor $1/d_n$	Lower percentage points						Upper percentage points					
		0·1	0·5	1·0	2·5	5·0	10·0	10·0	5·0	2·5	1·0	0·5	0·1
2	0·8862	0·00	0·01	0·02	0·04	0·09	0·18	2·33	2·77	3·17	3·64	3·97	4·65
3	·5908	0·06	0·13	0·19	0·30	0·43	0·62	2·90	3·31	3·68	4·12	4·42	5·06
4	·4857	0·20	0·34	0·43	0·59	0·76	0·98	3·24	3·63	3·98	4·40	4·69	5·31
5	·4299	0·37	0·55	0·66	0·85	1·03	1·26	3·48	3·86	4·20	4·60	4·89	5·48
6	0·3946	0·54	0·75	0·87	1·06	1·25	1·49	3·66	4·03	4·36	4·76	5·03	5·62
7	·3698	0·69	0·92	1·05	1·25	1·44	1·68	3·81	4·17	4·49	4·88	5·15	5·73
8	·3512	0·83	1·08	1·20	1·41	1·60	1·83	3·93	4·29	4·61	4·99	5·26	5·82
9	·3367	0·96	1·21	1·34	1·55	1·74	1·97	4·04	4·39	4·70	5·08	5·34	5·90
10	0·3249	1·08	1·33	1·47	1·67	1·86	2·09	4·13	4·47	4·79	5·16	5·42	5·97
11	·3152	1·20	1·45	1·58	1·78	1·97	2·20	4·21	4·55	4·86	5·23	5·49	6·04
12	·3069	1·30	1·55	1·68	1·88	2·07	2·30	4·29	4·62	4·92	5·29	5·54	6·09
13	·2998	1·39	1·64	1·77	1·97	2·16	2·39	4·35	4·68	4·99	5·35	5·60	6·14
14	·2935	1·47	1·72	1·86	2·06	2·24	2·47	4·41	4·74	5·04	5·40	5·65	6·19
15	0·2880	1·55	1·80	1·93	2·14	2·32	2·54	4·47	4·80	5·09	5·45	5·70	6·23
16	·2831	1·63	1·88	2·01	2·21	2·39	2·61	4·52	4·85	5·14	5·49	5·74	6·27
17	·2787	1·69	1·94	2·07	2·27	2·45	2·67	4·57	4·89	5·18	5·54	5·78	6·31
18	·2747	1·76	2·01	2·14	2·34	2·51	2·73	4·61	4·93	5·22	5·57	5·82	6·35
19	·2711	1·82	2·07	2·20	2·39	2·57	2·79	4·65	4·97	5·26	5·61	5·85	6·38
20	0·2677	1·87	2·12	2·25	2·45	2·62	2·84	4·69	5·01	5·30	5·65	5·89	6·41

The unit is the population standard deviation.
Estimate of σ = range (or mean range) in a sample of n observations $\times 1/d_n$.

This table is taken from the volume of Pearson and Hartley 1956, page 156. By permission. Biometrika Tables for Statisticians, Vol. 1, Cambridge Press, England.

BEST COPY AVAILABLE

TABLE 4.4 TEST STATISTIC $C_1 = \dfrac{(X_n - \bar{X})}{s_\nu}$

A. *Lower per cent points of the studentized extreme deviate* $(x_n - \bar{x})/s_\nu$ *or* $(\bar{x} - x_1)/s_\nu$

ν \ n	5 %							1 %						
	3	4	5	6	7	8	9	3	4	5	6	7	8	9
10	0·20	0·35	0·46	0·55	0·62	0·69	0·74	0·09	0·19	0·29	0·37	0·43	0·49	0·54
15	0·20	0·35	0·46	0·55	0·63	0·70	0·75	0·09	0·19	0·29	0·37	0·44	0·50	0·56
30	0·20	0·35	0·46	0·56	0·64	0·70	0·77	0·09	0·20	0·29	0·38	0·45	0·51	0·57
∞	0·20	0·35	0·47	0·56	0·65	0·72	0·78	0·09	0·20	0·30	0·38	0·46	0·53	0·59

B. *Upper per cent points of the studentized extreme deviate* $(x_n - \bar{x})/s_\nu$ *or* $(\bar{x} - x_1)/s_\nu$

ν \ n	5 %							1 %						
	3	4	5	6	7	8	9	3	4	5	6	7	8	9
10	2·02	2·29	2·49	2·63	2·75	2·85	2·93	2·76	3·05	3·25	3·39	3·50	3·59	3·67
11	1·99	2·26	2·44	2·58	2·70	2·79	2·87	2·71	3·00	3·19	3·33	3·44	3·53	3·61
12	1·97	2·22	2·40	2·54	2·65	2·75	2·83	2·67	2·95	3·14	3·28	3·39	3·48	3·55
13	1·95	2·20	2·38	2·51	2·62	2·71	2·79	2·63	2·91	3·10	3·24	3·34	3·43	3·51
14	1·93	2·18	2·35	2·48	2·59	2·68	2·76	2·60	2·87	3·06	3·20	3·30	3·39	3·47
15	1·92	2·16	2·33	2·46	2·56	2·65	2·73	2·57	2·84	3·02	3·16	3·27	3·35	3·43
16	1·90	2·14	2·31	2·44	2·54	2·63	2·70	2·55	2·81	3·00	3·13	3·24	3·32	3·39
17	1·89	2·13	2·30	2·42	2·52	2·61	2·68	2·52	2·79	2·97	3·10	3·21	3·29	3·36
18	1·88	2·12	2·28	2·41	2·51	2·59	2·66	2·50	2·77	2·95	3·08	3·18	3·27	3·34
19	1·87	2·11	2·27	2·39	2·49	2·58	2·65	2·49	2·75	2·92	3·06	3·16	3·24	3·31
20	1·87	2·10	2·26	2·38	2·48	2·56	2·63	2·47	2·73	2·91	3·04	3·14	3·22	3·29
24	1·84	2·07	2·23	2·35	2·44	2·52	2·59	2·43	2·68	2·85	2·97	3·07	3·15	3·22
30	1·82	2·04	2·20	2·31	2·40	2·48	2·55	2·38	2·62	2·79	2·91	3·01	3·08	3·15
40	1·80	2·02	2·17	2·28	2·37	2·44	2·51	2·34	2·57	2·73	2·85	2·94	3·02	3·08
60	1·78	1·99	2·14	2·25	2·33	2·41	2·47	2·30	2·52	2·68	2·79	2·88	2·95	3·01
120	1·76	1·97	2·11	2·21	2·30	2·37	2·43	2·25	2·48	2·62	2·73	2·82	2·89	2·95
∞	1·74	1·94	2·08	2·18	2·27	2·33	2·39	2·22	2·43	2·57	2·68	2·76	2·83	2·88

V is degrees of freedom of S_V

This table is taken from the paper of Nair 1948, page 143. By permission. Biometrika, Vol. 35.

BEST COPY AVAILABLE

TABLE 4.5. TEST STATISTIC $C_1 = \frac{(X_n - \bar{X})}{s_\nu}$

Critical Values for T When Standard Deviation s_ν is Independent of Present Sample

$$T' = \frac{x_n - \bar{x}}{s_\nu} \text{ or } \frac{\bar{x} - x_1}{s_\nu}$$

n	3	4	5	6	7	8	9	10	12
$\nu = df$	1% points								
10	2.78	3.10	3.32	3.48	3.62	3.73	3.82	3.90	4.04
11	2.72	3.02	3.24	3.39	3.52	3.63	3.72	3.79	3.93
12	2.67	2.96	3.17	3.32	3.45	3.55	3.64	3.71	3.84
13	2.63	2.92	3.12	3.27	3.38	3.48	3.57	3.64	3.76
14	2.60	2.88	3.07	3.22	3.33	3.43	3.51	3.58	3.70
15	2.57	2.84	3.03	3.17	3.29	3.38	3.46	3.53	3.65
16	2.54	2.81	3.00	3.14	3.25	3.34	3.42	3.49	3.60
17	2.52	2.79	2.97	3.11	3.22	3.31	3.38	3.45	3.56
18	2.50	2.77	2.95	3.08	3.19	3.28	3.35	3.42	3.53
19	2.49	2.75	2.93	3.06	3.16	3.25	3.33	3.39	3.50
20	2.47	2.73	2.91	3.04	3.14	3.23	3.30	3.37	3.47
24	2.42	2.68	2.84	2.97	3.07	3.16	3.23	3.29	3.38
30	2.38	2.62	2.79	2.91	3.01	3.08	3.15	3.21	3.30
40	2.34	2.57	2.73	2.85	2.94	3.02	3.08	3.13	3.22
60	2.29	2.52	2.68	2.79	2.88	2.95	3.01	3.06	3.15
120	2.25	2.48	2.62	2.73	2.82	2.89	2.95	3.00	3.08
∞	2.22	2.43	2.57	2.68	2.76	2.83	2.88	2.93	3.01
	5% points								
10	2.01	2.27	2.46	2.60	2.72	2.81	2.89	2.96	3.08
11	1.98	2.24	2.42	2.56	2.67	2.76	2.84	2.91	3.03
12	1.96	2.21	2.39	2.52	2.63	2.72	2.80	2.87	2.98
13	1.94	2.19	2.36	2.50	2.60	2.69	2.76	2.83	2.94
14	1.93	2.17	2.34	2.47	2.57	2.66	2.74	2.80	2.91
15	1.91	2.15	2.32	2.45	2.55	2.64	2.71	2.77	2.88
16	1.90	2.14	2.31	2.43	2.53	2.62	2.69	2.75	2.86
17	1.89	2.13	2.29	2.42	2.52	2.60	2.67	2.73	2.84
18	1.88	2.11	2.28	2.40	2.50	2.58	2.65	2.71	2.82
19	1.87	2.11	2.27	2.39	2.49	2.57	2.64	2.70	2.80
20	1.87	2.10	2.26	2.38	2.47	2.56	2.63	2.68	2.78
24	1.84	2.07	2.23	2.34	2.44	2.52	2.58	2.64	2.74
30	1.82	2.04	2.20	2.31	2.40	2.48	2.54	2.60	2.69
40	1.80	2.02	2.17	2.28	2.37	2.44	2.50	2.56	2.65
60	1.78	1.99	2.14	2.25	2.33	2.41	2.47	2.52	2.61
120	1.76	1.96	2.11	2.22	2.30	2.37	2.43	2.48	2.57
∞	1.74	1.94	2.08	2.18	2.27	2.33	2.39	2.44	2.52

The above percentage points are reproduced from H. A. David, "Revised upper percentage points of the extreme studentized deviate from the sample mean," Biometrika, Vol. 43 (1956), pp. 449–451. (Reference [3]).

This table is taken from Grubbs 1969, page 15.

BEST COPY AVAILABLE

TABLE 4.6A. TEST STATISTIC $C_3 = \frac{X_n - X_1}{s_\nu}$

Percentage points of the studentized range, $q = (x_n - x_1)/s_\nu$

Lower 5% points

ν \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
10	0·09	0·43	0·75	1·01	1·20	1·37	1·52	1·63	1·74	1·83	1·91	1·98	2·05	2·12	2·17	2·22	2·26	2·30	2·34
11	·09	·43	·75	·01	·21	·38	·52	·64	·75	·84	·92	2·00	·07	·13	·18	·24	·28	·33	·37
12	·09	·43	·75	·01	·21	·38	·53	·65	·76	·85	·93	·01	·08	·14	·20	·26	·30	·34	·38
13	·09	·43	·75	·01	·22	·39	·53	·65	·76	·86	·94	·02	·09	·15	·21	·27	·31	·36	·40
14	·09	·43	·75	·01	·22	·39	·54	·66	·77	·86	·95	·03	·10	·16	·22	·28	·32	·37	·41
15	0·09	0·43	0·75	1·01	1·22	1·39	1·54	1·66	1·77	1·87	1·95	2·03	2·11	2·17	2·23	2·29	2·34	2·38	2·43
16	·09	·43	·75	·01	·22	·39	·54	·67	·78	·87	·96	·04	·11	·18	·24	·30	·34	·39	·44
17	·09	·43	·75	·01	·22	·40	·55	·67	·78	·88	·97	·05	·12	·19	·25	·30	·35	·40	·45
18	·09	·43	·75	·02	·22	·40	·55	·67	·79	·88	·97	·05	·12	·19	·25	·31	·36	·41	·45
19	·09	·43	·75	·02	·23	·40	·55	·68	·79	·89	·98	·05	·13	·20	·26	·32	·37	·42	·46
20	0·09	0·43	0·75	1·02	1·23	1·40	1·55	1·68	1·79	1·89	1·98	2·06	2·13	2·20	2·27	2·32	2·37	2·42	2·47
24	·09	·43	·75	·02	·23	·41	·56	·69	·80	·90	·99	·08	·15	·22	·28	·34	·39	·45	·49
30	·09	·43	·76	·02	·24	·41	·57	·70	·81	·92	2·01	·09	·17	·24	·30	·36	·41	·47	·52
40	·09	·43	·76	·02	·24	·42	·57	·71	·82	·93	·02	·10	·18	·26	·32	·38	·43	·49	·54
60	0·09	0·43	0·76	1·02	1·24	1·43	1·58	1·72	1·83	1·94	2·04	2·12	2·20	2·28	2·34	2·40	2·46	2·52	2·57
120	·09	·43	·76	·03	·25	·43	·59	·73	·85	·96	·06	·14	·22	·30	·36	·43	·49	·54	·60
∞	0·09	0·43	0·76	1·03	1·25	1·44	1·60	1·74	1·86	1·97	2·07	2·16	2·24	2·32	2·39	2·45	2·52	2·57	2·62

Upper 5% points

ν \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18·0	27·0	32·8	37·1	40·4	43·1	45·4	47·4	49·1	50·6	52·0	53·2	54·3	55·4	56·3	57·2	58·0	58·8	59·6
2	6·09	8·3	9·8	10·9	11·7	12·4	13·0	13·5	14·0	14·4	14·7	15·1	15·4	15·7	15·9	16·1	16·4	16·6	16·8
3	4·50	5·91	6·82	7·50	8·04	8·48	8·85	9·18	9·46	9·72	9·95	10·15	10·35	10·52	10·69	10·84	10·98	11·11	11·24
4	3·93	5·04	5·76	6·29	6·71	7·05	7·35	7·60	7·83	8·03	8·21	8·37	8·52	8·66	8·79	8·91	9·03	9·13	9·23
5	3·64	4·60	5·22	5·67	6·03	6·33	6·58	6·80	6·99	7·17	7·32	7·47	7·60	7·72	7·83	7·93	8·03	8·12	8·21
6	3·46	4·34	4·90	5·31	5·63	5·89	6·12	6·32	6·49	6·65	6·79	6·92	7·03	7·14	7·24	7·34	7·43	7·51	7·59
7	3·34	4·16	4·63	5·06	5·36	5·61	5·82	6·00	6·16	6·30	6·43	6·55	6·66	6·76	6·85	6·94	7·02	7·09	7·17
8	3·26	4·04	4·53	4·89	5·17	5·40	5·60	5·77	5·92	6·05	6·18	6·29	6·39	6·48	6·57	6·65	6·73	6·80	6·87
9	3·20	3·95	4·42	4·76	5·02	5·24	5·43	5·60	5·74	5·87	5·98	6·09	6·19	6·28	6·36	6·44	6·51	6·58	6·64
10	3·15	3·88	4·33	4·65	4·91	5·12	5·30	5·46	5·60	5·72	5·83	5·93	6·03	6·11	6·20	6·27	6·34	6·40	6·47
11	3·11	3·82	4·26	4·57	4·82	5·03	5·20	5·35	5·49	5·61	5·71	5·81	5·90	5·99	6·06	6·14	6·20	6·26	6·33
12	3·08	3·77	4·20	4·51	4·75	4·95	5·12	5·27	5·40	5·51	5·62	5·71	5·80	5·88	5·95	6·03	6·09	6·15	6·21
13	3·06	3·73	4·15	4·45	4·69	4·88	5·05	5·19	5·32	5·43	5·53	5·63	5·71	5·79	5·86	5·93	6·00	6·05	6·11
14	3·03	3·70	4·11	4·41	4·64	4·83	4·99	5·13	5·25	5·36	5·46	5·55	5·64	5·72	5·79	5·85	5·92	5·97	6·03
15	3·01	3·67	4·08	4·37	4·60	4·78	4·94	5·08	5·20	5·31	5·40	5·49	5·58	5·65	5·72	5·79	5·85	5·90	5·96
16	3·00	3·65	4·05	4·33	4·56	4·74	4·90	5·03	5·15	5·26	5·35	5·44	5·52	5·59	5·66	5·72	5·79	5·84	5·90
17	2·98	3·63	4·02	4·30	4·52	4·71	4·86	4·99	5·11	5·21	5·31	5·39	5·47	5·55	5·61	5·68	5·74	5·79	5·84
18	2·97	3·61	4·00	4·28	4·49	4·67	4·82	4·96	5·07	5·17	5·27	5·35	5·43	5·50	5·57	5·63	5·69	5·74	5·79
19	2·96	3·59	3·98	4·25	4·47	4·65	4·79	4·92	5·04	5·14	5·23	5·32	5·39	5·46	5·53	5·59	5·65	5·70	5·75
20	2·95	3·58	3·96	4·23	4·45	4·62	4·77	4·90	5·01	5·11	5·20	5·28	5·36	5·43	5·49	5·55	5·61	5·66	5·71
24	2·92	3·53	3·90	4·17	4·37	4·54	4·68	4·81	4·92	5·01	5·10	5·18	5·25	5·32	5·38	5·44	5·50	5·54	5·59
30	2·89	3·49	3·84	4·10	4·30	4·46	4·60	4·72	4·83	4·92	5·00	5·08	5·15	5·21	5·27	5·33	5·38	5·43	5·48
40	2·86	3·44	3·79	4·04	4·23	4·39	4·52	4·63	4·74	4·82	4·91	4·98	5·05	5·11	5·16	5·22	5·27	5·31	5·36
60	2·83	3·40	3·74	3·98	4·16	4·31	4·44	4·55	4·65	4·73	4·81	4·88	4·94	5·00	5·06	5·11	5·16	5·20	5·24
120	2·80	3·36	3·69	3·92	4·10	4·24	4·36	4·48	4·56	4·64	4·72	4·78	4·84	4·90	4·95	5·00	5·05	5·09	5·13
∞	2·77	3·31	3·63	3·86	4·03	4·17	4·29	4·39	4·47	4·55	4·62	4·68	4·74	4·80	4·85	4·89	4·93	4·97	5·01

n is the size of sample from which the range is obtained and ν is the number of degrees of freedom of s_ν.

This table is taken from Pearson and Hartley 1956, page 176. By permission.

BEST COPY AVAILABLE

TABLE 4.6B. TEST STATISTIC $C_3 = \frac{X_n - X_1}{s_\nu}$

Lower 1% points

ν \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
10	0·02	0·18	0·42	0·64	0·81	0·96	1·11	1·23	1·34	1·41	1·50	1·57	1·62	1·70	1·74	1·81	1·84	1·88	1·92
11	·02	·18	·42	·64	·82	·97	·12	·24	·35	·43	·52	·58	·64	·71	·76	·82	·86	·91	·94
12	·02	·18	·42	·64	·82	·98	·12	·24	·35	·44	·53	·60	·65	·73	·77	·84	·88	·92	·96
13	·02	·18	·42	·64	·83	·98	·13	·25	·36	·45	·54	·61	·66	·74	·79	·85	·89	·94	·98
14	·02	·18	·42	·65	·83	·99	·13	·25	·37	·46	·55	·62	·68	·76	·80	·87	·91	·95	2·00
15	0·02	0·18	0·42	0·65	0·83	0·99	1·14	1·26	1·37	1·46	1·55	1·63	1·69	1·76	1·81	1·88	1·92	1·97	2·01
16	·02	·18	·42	·65	·83	·99	·14	·26	·37	·47	·56	·63	·70	·77	·82	·89	·93	·98	·02
17	·02	·18	·42	·65	·84	1·00	·14	·27	·38	·48	·57	·64	·70	·78	·83	·90	·94	·99	·04
18	·02	·18	·42	·65	·84	·00	·15	·27	·38	·48	·57	·65	·71	·79	·84	·91	·95	2·00	·05
19	·02	·18	·43	·65	·84	·00	·15	·28	·39	·48	·58	·65	·72	·80	·85	·91	·96	·01	·06
20	0·02	0·18	0·43	0·65	0·84	1·01	1·15	1·28	1·39	1·49	1·58	1·66	1·72	1·80	1·85	1·92	1·97	2·01	2·06
24	·02	·18	·43	·65	·85	·01	·16	·29	·40	·50	·60	·67	·74	·82	·88	·94	·99	·05	·09
30	·02	·18	·43	·66	·85	·02	·17	·30	·41	·52	·61	·69	·76	·84	·90	·97	2·02	·08	·12
40	·02	·18	·43	·66	·85	·02	·18	·31	·43	·53	·63	·71	·79	·86	·92	·99	·04	·10	·15
60	0·02	0·18	0·43	0·66	0·86	1·03	1·19	1·32	1·44	1·55	1·64	1·73	1·81	1·88	1·95	2·02	2·07	2·13	2·18
120	·02	·18	·43	·66	·86	·04	·20	·33	·45	·56	·66	·75	·83	·91	·98	·04	·10	·16	·21
∞	0·02	0·19	0·43	0·66	0·87	1·05	1·20	1·34	1·47	1·58	1·68	1·77	1·86	1·93	2·01	2·08	2·14	2·20	2·25

Upper 1% points

ν \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	90·0	135	164	186	202	216	227	237	246	253	260	266	272	277	282	286	290	294	298
2	14·0	19·0	22·3	24·7	26·6	28·2	29·5	30·7	31·7	32·6	33·4	34·1	34·8	35·4	36·0	36·5	37·0	37·5	37·9
3	8·26	10·6	12·2	13·3	14·2	15·0	15·6	16·2	16·7	17·1	17·5	17·9	18·2	18·5	18·8	19·1	19·3	19·5	19·8
4	6·51	8·12	9·17	9·96	10·6	11·1	11·5	11·9	12·3	12·6	12·8	13·1	13·3	13·5	13·7	13·9	14·1	14·2	14·4
5	5·70	6·97	7·80	8·42	8·91	9·32	9·67	9·97	10·24	10·48	10·70	10·89	11·08	11·24	11·40	11·55	11·68	11·81	11·93
6	5·24	6·33	7·03	7·56	7·97	8·32	8·61	8·87	9·10	9·30	9·49	9·65	9·81	9·95	10·08	10·21	10·32	10·43	10·54
7	4·95	5·92	6·54	7·01	7·37	7·68	7·94	8·17	8·37	8·55	8·71	8·86	9·00	9·12	9·24	9·35	9·46	9·55	9·65
8	4·74	5·63	6·20	6·63	6·96	7·24	7·47	7·68	7·87	8·03	8·18	8·31	8·44	8·55	8·66	8·76	8·85	8·94	9·03
9	4·60	5·43	5·96	6·35	6·66	6·91	7·13	7·32	7·49	7·65	7·78	7·91	8·03	8·13	8·23	8·32	8·41	8·49	8·57
10	4·48	5·27	5·77	6·14	6·43	6·67	6·87	7·05	7·21	7·36	7·48	7·60	7·71	7·81	7·91	7·99	8·07	8·15	8·22
11	4·39	5·14	5·62	5·97	6·25	6·48	6·67	6·84	6·99	7·13	7·25	7·36	7·46	7·56	7·65	7·73	7·81	7·88	7·95
12	4·32	5·04	5·50	5·84	6·10	6·32	6·51	6·67	6·81	6·94	7·06	7·17	7·26	7·36	7·44	7·52	7·59	7·66	7·73
13	4·26	4·96	5·40	5·73	5·98	6·19	6·37	6·53	6·67	6·79	6·90	7·01	7·10	7·19	7·27	7·34	7·42	7·48	7·55
14	4·21	4·89	5·32	5·63	5·88	6·08	6·26	6·41	6·54	6·66	6·77	6·87	6·96	7·05	7·12	7·20	7·27	7·33	7·39
15	4·17	4·83	5·25	5·53	5·80	5·99	6·16	6·31	6·44	6·55	6·66	6·76	6·84	6·93	7·00	7·07	7·14	7·20	7·26
16	4·13	4·78	5·19	5·49	5·72	5·92	6·08	6·22	6·35	6·46	6·56	6·66	6·74	6·82	6·90	6·97	7·03	7·09	7·15
17	4·10	4·74	5·14	5·43	5·66	5·85	6·01	6·15	6·27	6·38	6·48	6·57	6·66	6·73	6·80	6·87	6·94	7·00	7·05
18	4·07	4·70	5·09	5·38	5·60	5·79	5·94	6·08	6·20	6·31	6·41	6·50	6·58	6·65	6·72	6·79	6·85	6·91	6·96
19	4·05	4·67	5·05	5·33	5·55	5·73	5·89	6·02	6·14	6·25	6·34	6·43	6·51	6·58	6·65	6·72	6·78	6·84	6·89
20	4·02	4·64	5·02	5·29	5·51	5·69	5·84	5·97	6·09	6·19	6·29	6·37	6·45	6·52	6·59	6·65	6·71	6·76	6·82
24	3·96	4·54	4·91	5·17	5·37	5·54	5·69	5·81	5·92	6·02	6·11	6·19	6·26	6·33	6·39	6·45	6·51	6·56	6·61
30	3·89	4·45	4·80	5·05	5·24	5·40	5·54	5·65	5·76	5·85	5·93	6·01	6·08	6·14	6·20	6·26	6·31	6·36	6·41
40	3·82	4·37	4·70	4·93	5·11	5·27	5·39	5·50	5·60	5·69	5·77	5·84	5·90	5·96	6·02	6·07	6·12	6·17	6·21
60	3·76	4·28	4·60	4·82	4·99	5·13	5·25	5·36	5·45	5·53	5·60	5·67	5·73	5·79	5·84	5·89	5·93	5·98	6·02
120	3·70	4·20	4·50	4·71	4·87	5·01	5·12	5·21	5·30	5·38	5·44	5·51	5·56	5·61	5·66	5·71	5·75	5·79	5·83
∞	3·64	4·12	4·40	4·60	4·76	4·88	4·99	5·08	5·16	5·23	5·29	5·35	5·40	5·45	5·49	5·54	5·57	5·61	5·65

n is the size of the sample from which the range is obtained and ν is the number of degrees of freedom of s_ν.

This table is taken from Pearson and Hartley 1956, page 177. By permission.

BEST COPY AVAILABLE

TABLE 4.7. TEST STATISTIC $D_1 = \frac{X_n - \bar{X}}{s}$ or $\frac{\bar{X} - X_1}{s}$

Number of Observations n	5% Significance Level	2.5% Significance Level	1% Significance Level
3	1.15	1.15	1.15
4	1.46	1.48	1.49
5	1.67	1.71	1.75
6	1.82	1.89	1.94
7	1.94	2.02	2.10
8	2.03	2.13	2.22
9	2.11	2.21	2.32
10	2.18	2.29	2.41
11	2.23	2.36	2.48
12	2.29	2.41	2.55
13	2.33	2.46	2.61
14	2.37	2.51	2.66
15	2.41	2.55	2.71
16	2.44	2.59	2.75
17	2.47	2.62	2.79
18	2.50	2.65	2.82
19	2.53	2.68	2.85
20	2.56	2.71	2.88
21	2.58	2.73	2.91
22	2.60	2.76	2.94
23	2.62	2.78	2.96
24	2.64	2.80	2.99
25	2.66	2.82	3.01
30	2.75	2.91	
35	2.82	2.98	
40	2.87	3.04	
45	2.92	3.09	
50	2.96	3.13	
60	3.03	3.20	
70	3.09	3.26	
80	3.14	3.31	
90	3.18	3.35	
100	3.21	3.38	

$$D_{1n} = \frac{x_n - \bar{x}}{s} \qquad s = \left\{\frac{\sum (x_i - \bar{x})^2}{n-1}\right\}^{\frac{1}{2}} = \left\{\frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)}\right\}^{\frac{1}{2}}$$

$$D_{1n} = \frac{\bar{x} - x_1}{s} \qquad x_1 \leq x_2 \leq \cdots \leq x_n$$

Note: Values of D for $n \leq 25$ are based on those given in Reference [8]. For $n > 25$, the values of D are approximated. All values have been adjusted for division by $n - 1$ instead of n in calculating s.

This table is taken from Grubbs 1969, page 4.

BEST COPY AVAILABLE

TABLE 4.8. TEST STATISTIC $D_3 = \frac{\text{Range (w)}}{s}$ or $\frac{X_n - X_1}{s}$

*Critical Values for w/s (Ratio of Range to Sample Standard Deviation)**

Number of Observations n	5% Significance Level	1% Significance Level	0.5% Significance Level
3	2.00	2.00	2.00
4	2.43	2.44	2.45
5	2.75	2.80	2.81
6	3.01	3.10	3.12
7	3.22	3.34	3.37
8	3.40	3.54	3.58
9	3.55	3.72	3.77
10	3.68	3.88	3.94
11	3.80	4.01	4.08
12	3.91	4.13	4.21
13	4.00	4.24	4.32
14	4.09	4.34	4.43
15	4.17	4.43	4.53
16	4.24	4.51	4.62
17	4.31	4.59	4.69
18	4.38	4.66	4.77
19	4.43	4.73	4.84
20	4.49	4.79	4.91
30	4.89	5.25	5.39
40	5.15	5.54	5.69
50	5.35	5.77	5.91
60	5.50	5.93	6.09
80	5.73	6.18	6.35
100	5.90	6.36	6.54
150	6.18	6.64	6.84
200	6.38	6.85	7.03
500	6.94	7.42	7.60
1000	7.33	7.80	7.99

* Taken from H. A. David, H. O. Hartley and E. S. Pearson, "The Distribution of the Ratio in a Single Sample of Range to Standard Deviation," Biometrika, Vol. 41 (1954), pp. 482–493. (Reference [4])

$w = x_n - x_1$

$x_1 \leq x_2 \leq \cdots \leq x_n$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

This table is taken from Grubbs 1969, page 8.

BEST COPY AVAILABLE

TABLE 4.9. DIXON RATIO TEST STATISTICS

CRITICAL VALUES AND CRITERIA FOR TESTING FOR EXTREME VALUES

N \ α	.30	.20	.10	.05	.02	.01	.005	Criterion
3	.684	.781	.886	.941	.976	.988	.994	
4	.471	.560	.679	.765	.846	.889	.926	
5	.373	.451	.557	.642	.729	.780	.821	$r_{10} = \frac{x_N - x_{N-1}}{x_N - x_1}$
6	.318	.386	.482	.560	.644	.698	.740	
7	.281	.344	.434	.507	.586	.637	.680	
8	.318	.385	.479	.554	.631	.683	.725	
9	.288	.352	.441	.512	.587	.635	.677	$r_{11} = \frac{x_N - x_{N-1}}{x_N - x_2}$
10	.265	.325	.409	.477	.551	.597	.639	
11	.391	.442	.517	.576	.638	.679	.713	
12	.370	.419	.490	.546	.605	.642	.675	$r_{21} = \frac{x_N - x_{N-2}}{x_N - x_2}$
13	.351	.399	.467	.521	.578	.615	.649	
14	.370	.421	.492	.546	.602	.641	.674	
15	.353	.402	.472	.525	.579	.616	.647	
16	.338	.386	.454	.507	.559	.595	.624	
17	.325	.373	.438	.490	.542	.577	.605	
18	.314	.361	.424	.475	.527	.561	.589	
19	.304	.350	.412	.462	.514	.547	.575	
20	.295	.340	.401	.450	.502	.535	.562	$r_{22} = \frac{x_N - x_{N-2}}{x_N - x_3}$
21	.287	.331	.391	.440	.491	.524	.551	
22	.280	.323	.382	.430	.481	.514	.541	
23	.274	.316	.374	.421	.472	.505	.532	
24	.268	.310	.367	.413	.464	.497	.524	
25	.262	.304	.360	.406	.457	.489	.516	

Reproduced from: W. J. Dixon, "Processing Data for Outliers." Biometrics 9:74–89, 1953. With permission of the Biometric Society.

V. SUMMARY

This outline discussed the various test statistics available to test a single suspected outlying observation. The criteria for selection of a given test depends upon the quality of knowledge known about the population (μ and σ); the power or efficiency of the test statistic; the ruggedness against masking by possible multiple outliers; and the ease of computation. Tables of critical values are appended for the recommended tests.

BEST COPY AVAILABLE

REFERENCES

1. David, H. A., *Order Statistics*, John Wiley & Sons, Inc. New York, N.Y. 1970.

2. Dixon, W. J., Analysis of Extreme Values, Ann. Math. Statist. 21, 488—506, 1950.

3. Dixon, W. J., Processing Data for Outliers, Biometrics 9, 74—89, 1953.

4. Grubbs, F. E., Sample Criteria for Testing Outlying Observations, Ann. Math. Statist. 21, 27—58, 1950.

5. Grubbs, F. E., Procedures for Detecting Outlying Observations in Samples, Technometrics, 11, 1—21, 1969.

6. Grubbs, F. E., Procedures for Detecting Outlying Observations, Chapter 3 of Army Statistics Manual DARCOM—P706-103, U.S. Army Aberdeen Research and Development Center, Aberdeen Proving Ground, MD 21005, In Press 1979.

7. Irwin, J. O., On a Criterion for the Rejection of Outlying Observations, Biometrika 17, 238—50, 1925.

8. Nair, K. R., The Distribution of the Extreme Deviate from the Sample Mean and Its Studentized Form, Biometrika 35, 118—44, 1948.

9. Pearson, E. S. and Hartley, H. O. (editors) Biometrika Tables for Statisticians Vol. 1, Cambridge Press, England 1956.

10. "Student," Errors of Routine Analysis, Biometrika, 19, 151—164, 1927.

BIBLIOGRAPHY

1. Ferguson, T. S., On the Rejection of Outliers, Proc. 4th Berkeley Symposium I, 253—87, 1961.

2. McKay, A. T. The Distribution of the Difference Between the Extreme Observation and the Sample Mean in Samples of n From a Normal Universe, Biometrika, 27, 466—71, 1935.

BEST COPY AVAILABLE

BLANK PAGE

Lesson 5

EXAMPLES AND EXERCISES – SINGLE OUTLIERS

TIME: Fifth Period, First Day (150 minutes)

OBJECTIVES: Enable the trainee to complete a set of examples and problems so that he can exercise judgment as to inclusion of a single outlier regardless of its position in the data set.

EXHIBITS: 5.1 through 5.8, Critical Value Diagrams

KEY TERMS: No new key terms

MATERIALS: Student Manual

Pencils with Erasers

Calculators with Log Functions

Overhead Projector w/Screen

Transparencies 5.1 through 5.2

BEST COPY AVAILABLE

ABSTRACT

This lesson provides examples of how to treat outliers under various conditions of prior knowledge about the data set. It also provides the opportunity to practice the treatment techniques learned. The following is an outline of the Examples and Exercises lesson:

- Introduction
- Examples
- Exercises
- Solutions to Examples and Exercises

BEST COPY AVAILABLE

Lesson 5

EXAMPLES AND EXERCISES — SINGLE OUTLIERS

INTRODUCTION

In Lesson 4 you learned about the use of a number of statistical tests together with their associated tables which enabled you to make a statistical decision regarding the use of an outlier under various conditions of prior knowledge about the data set.

The following examples and exercises will give you practice using these techniques, as well as give you an opportunity to develop further your computational skills.

1. Work by yourself unless the instructor directs otherwise.
2. Refer to Lesson 4 as the need arises.
3. Ask the instructor for help as you need it. Be prepared to discuss your solutions with the rest of the class.

EXERCISES

Complete the following exercises according to the directions for each one.

1. This problem uses the same data as for Problem 1 of Lesson 3. However, in this case, the population mean, μ, is unknown.

   Using Test $B1 = \frac{(Xn - \overline{X})}{\sigma}$ and $\sigma = 2$:

   a. Calculate the mean $\overline{X}$

   b. Determine whether or not any of the data points should be rejected.

2. This problem uses the same data as for Problem 2 of Lesson 3. However the population mean is unknown.

   Using Test $B1 = \frac{(\overline{X} - X_1)}{\sigma}$ and $\sigma = 20$:

   a. Calculate the mean $\overline{X}$

   b. Determine whether the lowest value 58, should be rejected.

3. This problem uses the same data as for Problem 3 of Lesson 3. However, the population mean is not known.

   Using test $B1 = \frac{(\overline{X} - X_1)}{\sigma}$ and $\sigma = 10$:

   a. Calculate the mean $\overline{X}$

   b. Determine whether the value 21 should be rejected.

4. This problem uses the same data as for Problem 1 of Lesson 3.

   Using Test $B4 = \frac{\text{Range}}{\sigma}$ and $\sigma = 2$:

   Determine whether a value should be rejected.

5. This problem uses the same data as for Problem 2 of Lesson 3.

   Using Test $B4 = \frac{\text{Range}}{\sigma}$ and $\sigma = 20$:

   Determine whether a value should be rejected.

BEST COPY AVAILABLE

6. This problem uses the same data as for Problem 3 of Lesson 3.

   Using Test B4 $= \frac{\text{Range}}{\sigma}$ and $\sigma = 10$:

   Determine whether a value should be rejected.

7. An assistant reported the following set of data for carbon monoxide detector tube readings.

   48, 51, 47, 49, 50, 52, 50, 51, 49, 50

   Is the data from a normal population with a σ = 10?

   Using test B4 $= \frac{\text{Range}}{\sigma}$ and $\sigma = 10$:

   Determine if this sample represents the data to be expected from such a series of measurements.

8. Ten measurements of the concentration of a nuisance respirable dust were made in a plant. The calculated values are: 8.78 mg/m^3, 9.87, 9.24, 12.29, 9.77, 12.67, 9.50, 7.96, 8.44 and 8.63. Twenty similar measurements give an expected standard deviation of 2.66 mg/m^3. Calculate the mean $\bar{X}$.

   Using Test C1 $= \frac{(X_n - \bar{X})}{s_v}$ or $\frac{(\bar{X} - X_1)}{s_v}$:

   Determine whether a value should be rejected.

9. Ten respirable dust measurements were taken. The values were 7.55 mg/m^3, 9.74, 8.48, 14.58, 9.53, 15.33, 9.00, 5.91, 6.88, and 7.25. Twenty similar measurements give a s_v of 2.66 mg/m^3. Calculate the mean $\bar{X}$:

   Using test C1 $= \frac{(X_n - \bar{X})}{s_v}$ or $\frac{(\bar{X} - X_1)}{s_v}$

   Determine whether a value should be rejected.

10. Ten total dust measurements were taken. The calculated values are: 22.08 mg/m^3, 18.87, 19.57, 21.80, 19.73, 17.83, 21.74, 19.88, 21.07, and 20.51. s_v = 2.66 Calculate the mean $\bar{X}$:

    Using test C1 $= \frac{(X_n - \bar{X})}{s_v}$ or $\frac{(\bar{X} - X_1)}{s_v}$

    Determine whether a value should be rejected.

11. Ten total dust measurements were made. The calculated values were 11.57 mg/m^3, 8.36, 9.06, 11.29, 9.27, 7.32, 11.23, 9.37, 10.56 and 10.00. Similar measurements give a s_v of 0.89. Calculate the mean $\bar{X}$:

    Using test C1 $= \frac{(X_n - \bar{X})}{s_v}$ or $\frac{(\bar{X} - X_1)}{s_v}$:

    Determine whether a value should be rejected.

BEST COPY AVAILABLE

12. Using the data from problems 8, 9, 10, and 11:

Problem 8	*Problem 9*	*Problem 10*	*Problem 11*
8.8	7.5	22.1	11.6
9.9	9.7	18.9	8.4
9.2	8.5	19.6	9.1
12.3	14.6	21.8	11.3
9.8	9.5	19.7	9.2
12.7	15.3	17.8	7.3
9.5	9.0	21.7	11.2
8.0	5.9	19.9	9.4
8.4	6.9	21.1	10.5
8.6	7.3	20.5	10.0
$s_{r19} = 2.66$	2.66	2.66	$s_r = .89$

Using test statistic $C3 = \frac{\text{range}}{s_r}$

Compare the decisions on rejection with those made in problems 8, 9, 10, and 11.

13. Use the data from Problem 1 of Lesson 3 without prior knowledge of the population mean or standard deviation.

    Use the test statistic $D_1 = \frac{X_n - \overline{X}}{s}$:

    a. Calculate the mean, $\overline{X}$

    b. Calculate the standard deviation, s

    c. Determine whether any of the data points should be rejected.

14. Use the data from Problem 2 of Lesson 3 without prior knowledge of population mean or standard deviation.

    Use the test statistic $D_1 = \frac{X_n - \overline{X}}{s}$:

    a. Calculate the mean $\overline{X}$

    b. Calculate the standard deviation, s

    c. Determine whether any of the data points should be rejected.

15. Using the data from Problem 3 of Lesson 3 without prior knowledge of the population mean or standard deviation.

    Use the test statistic $D_1 = \frac{X_n - \overline{X}}{s}$:

    a. Calculate the mean $\overline{X}$

    b. Calculate the standard deviation, s

    c. Determine whether any of the data points should be rejected.

16. Apply test statistic D_3 to the data set from Problem 13 of Lesson 5. Use:

    $D_3 = \frac{X_n - X_1}{s}$ or $\frac{\text{Range (W)}}{s}$

BEST COPY AVAILABLE

17. Apply test statistic D_3 to the data set from Problem 14 of Lesson 5.

18. Apply test statistic D_3 to the data set from Problem 15 of Lesson 5.

19. Apply the Dixon test statistic to the data set of Problem 13 of Lesson 5.

$n = 20$ Use r_{22} from table 9

$$r_{22} = \frac{Xn - X_{n-2}}{Xn - X_3}$$

Arrange Data Set in Ascending Order of Values

5.6, 5.8, 6.6, 6.6, 7.2, 8.1, 8.2, 8.6, 8.9, 9.7, 9.9, 10.1, 10.8, 10.9, 11.0, 11.1, 12.7, 13.4, 13.6, 14.6.

20. Apply the Dixon test statistic to the data set of Problem 14 of Lesson 5.

Arrange data in ascending order — 58, 78, 87, 92, 93, 97, 100, 103, 106, 123

$N = 10$ so use Table r_{11}

$$r_{11} = \frac{X_n - X_{n-1}}{X_n - X_2}$$

21. Apply the Dixon test statistic to the following set of data.

44, 21, 50, 46, 50, 46, 49, 40, 58, 43

Put data in ascending order

21, 40, 43, 44, 46, 46, 49, 50, 50, 58

22. Apply the Dixon test statistic to the data set from Problem 15 of Lesson 5.

List data in ascending order — 21, 42, 43, 44, 46, 46, 50, 50, 66

$$r_{11} = \frac{X_n - X_{n-1}}{X_n - X_2}$$

BEST COPY AVAILABLE

Lesson 6

MULTIPLE OUTLIERS

TIME: Sixth Period, Second Day

OBJECTIVES:

1. Enable the trainee to understand how to treat the data when:
   a. Outliers are split high and low
   b. Multiple outliers appear on either the high or low side of the data distribution.
2. Familiarize the trainee with non-parametric tests, the Slippage Test.

EXHIBITS:

6.1 Critical Values for $\frac{W}{s}$

6.2 Critical Values for S^2_{n-1}, $\frac{n}{S^2}$ or $\frac{S^2_{1,2}}{S^2}$ for simultaneously testing the two largest or two smallest observations

6.3 Percentage Points of Rosner's RST Many Outliers Test Statistics

6.4 Percentage Points of Hawkins E_k

6.5 Slippage Test Data Table

6.6 Quantiles of the K-sample Slippage Test Statistic

6.7 Slippage Test Statistic Asbestos Measurement Values

6.8 Approximate Five Percent Two-Tail Limits for Ranking Scores

Appendix Table A — Table of Critical Values for S^2_{n-1}, $\frac{n}{S^2}$, $\frac{S^2_{1,2}}{S^2}$ for Simultaneously Testing the Two Largest or Two Smallest Observations

KEY TERMS:

Slippage Tests
Extreme Rank Sum Tests
Trimmed Mean
Trimmed Variance

MATERIALS:

Chalkboard and Chalk
Overhead Projector and Screen
Exhibits 6.1 -- 6.8

METHODS: Explanation/Lecture

BEST COPY AVAILABLE

ABSTRACT

The following is an outline of the contents of the Multiple Outliers lesson:

- Introduction
- One High and One Low Outlier
- Two High or Two Low Outliers
- Rejection of Multiple Outliers
- Slippage Tests
- Youden Extreme Rank Sum Test
- Summary
- References
- Appendix

BEST COPY AVAILABLE

Lesson 6

MULTIPLE OUTLIERS

INTRODUCTION

Two previous lessons treated the general philosophical problem of outlying observations and the approaches to dealing with only one outlier per sample. The approaches to dealing with a single outlier were modified depending upon the available amount of independent knowledge about the population variance, σ^2. In dealing with multiple outliers in a single set of data most applications deal with no prior knowledge of μ or σ — the "D" type statistic.

In this outline, we will first treat the case of handling one high and one low outlier; two high or two low outliers; multiple outliers; and a special non-parametric test, the slippage type test.

EXPLANATION/LECTURE

I. ONE HIGH AND ONE LOW OUTLIER

The range type test is recommended because it is designed to handle upper and lower outliers in contrast to the case of the outliers being only on one side of the distribution. The test statistic of David, Hartley and Pearson 1966 (12 lesson 004.22) has been chosen for ASTM E-11 use by Grubbs 1969.[3]

The test is: $R = \frac{w}{s}$

Where $w = Xn - X_1$ and $s = \sqrt{\frac{\Sigma (X_i - \bar{X})^2}{n-1}}$

The classical case to which, it seems, all tests are applied are the set of 15 observations of the "vertical semi-diameters of Venus" made by a Lieutenant Herndon in 1846 and treated by Chauvenet in 1876. The residuals in X_1, X_2, etc. order are:

−1.40	−0.24	−0.05	0.18	0.48
−0.44	−0.22	0.06	0.20	0.63
−0.33	−0.13	0.10	0.39	1.01

Using the above set of data one can calculate that $w = 1.01 - (-1.40) = 2.41$ and $s = 0.551$ as $\bar{X} = 0.018$. The test statistic or ratio

$$R = \frac{w}{s} = \frac{2.41}{0.551} = 4.37$$

The Table of critical values for this test table — (the same table as for test D_3) is entered for an n of 15. The test statistic value of 4.374 falls between the critical values for the 5% (4.17) and the 1% (4.43). If both the X_1 and X_n were about equidistant from the mean one could conclude that both values were outliers. Since, however, this is not true t-statistic type test such as the

$$D_1 \text{ test } D_1 = \frac{X_n - \bar{X}}{s} \text{ or } \frac{\bar{X} - X_1}{s}$$

should be used to test the largest suspect outlier observation −1.40. Applying the test D_1, reject the value −1.40. A second application of test D_1 to the other suspect outlier,

BEST COPY AVAILABLE

TABLE 6.1. CRITICAL VALUES FOR W/S (RATIO OF RANGE TO SAMPLE STANDARD DEVIATION)*

Number of Observations n	5% Significance Level	1% Significance Level	0.5% Significance Level
3	2.00	2.00	2.00
4	2.43	2.44	2.45
5	2.75	2.80	2.81
6	3.01	3.10	3.12
7	3.22	3.34	3.37
8	3.40	3.54	3.58
9	3.55	3.72	3.77
10	3.68	3.88	3.94
11	3.80	4.01	4.08
12	3.91	4.13	4.21
13	4.00	4.24	4.32
14	4.09	4.34	4.43
15	4.17	4.43	4.53
16	4.24	4.51	4.62
17	4.31	4.59	4.69
18	4.38	4.66	4.77
19	4.43	4.73	4.84
20	4.49	4.79	4.91
30	4.89	5.25	5.39
40	5.15	5.54	5.69
50	5.35	5.77	5.91
60	5.50	5.93	6.09
80	5.73	6.18	6.35
100	5.90	6.36	6.54
150	6.18	6.64	6.84
200	6.38	6.85	7.03
500	6.94	7.42	7.60
1000	7.33	7.80	7.99

*Taken from H. A. David, H. O. Hartley and E. S. Pearson, "The Distribution of the Ratio of Range to Standard Deviation," Biometrika, Vol. 41 (1954), pp. 482—493. (Reference (4)). By permission.

$W = X_n - X_1$

$X_1 \leq X_2 \ldots \leq X_n$

$$s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n-1}}$$

BEST COPY AVAILABLE

1.01 against the new total of 14 observations fails to reject the value. The Dixon ratio test could also be applied to these confirmatory tests as well and will yield equivalent decisions.

II. TWO HIGH OR TWO LOW OUTLIERS

This case has been well documented by Grubbs in 1950,[2] 1969,[3] and Grubbs and Beck 1972.[5] McMillan and David in 1971 have also studied the case when the variance is known and McMillan 1971,[6] and Moran and McMillan 1973,[8] with unknown variance. The Grubbs test was mentioned in the second lesson as test statistic D_6.

The test statistic D_6 is abbreviated as

$$\frac{S^2_{n-1,n}}{S^2} \text{ or } \frac{S^2_{1,2}}{S^2} \text{ where } S^2 = \sum_{i=1}^{n} (Xi - \overline{X})^2$$

$$\text{where } \overline{X} = \frac{1}{n}\sum_{i=1}^{n} Xi, \; S^2_{n-1,n} = \sum_{i=1}^{n-2} \frac{(Xi - \overline{X}_{n-1,n})^2}{n-2}$$

$$\text{where } \overline{X}_{n-1,n} = \frac{1}{n-2}\sum_{i=1}^{n-2} Xi \quad S^2_{1,2} = \sum_{i=3}^{n}$$

$$\frac{(Xi - \overline{X}_{1,2})^2}{n-2} \qquad \text{where } \overline{X}_{1,2} = \frac{1}{n-2}\sum_{i=3}^{n} Xi$$

Table 2 gives the Critical Values for the test statistic D_6. The example set of data for this test consists of ten observations, put in ascending order they are: 2.02, 2.22, 3.04, 3.23, 3.59, 3.73, 3.94, 4.05, 4.11 and 4.13. Applying the test statistic $\frac{S_{1,2}}{S^2}$ would determine whether the values 2.02 and 2.22 should be rejected. Using the above formulas we calculate:

$$S^2 = \sum_{i=1}^{n} (Xi - \overline{X})^2 = \frac{n\sum X_i^2 - (\sum Xi)^2}{n}$$

$$= \frac{10(121.3594) - (34.06)^2}{10} \text{ or } S^2 = 5.351$$

$$S^2_{1,2} = \sum_{i=3}^{n} (Xi - X_{1,2})^2$$

$$= \frac{(n-2)\sum_{i=1}^{n} X_i^2 - (\sum_{i=3}^{n} X_i)^2}{n-2}$$

$$= \frac{8(112.3506) - (29.82)^2}{8} = \frac{9.5724}{8} = 1.197$$

$$D_6 = \frac{S^2_{1,2}}{S^2} = \frac{1.197}{5.351} = 0.224$$

Comparing the $D_6 = 0.224$ to the values in Table 2 for an $n = 10$ the 10% Critical Value (CV) is 0.2853; the 5% CV is 0.2305 and the 1% CV is 0.1415. The test value $D_6 = 0.224$ falls between the 5% and 1% C.V.'s. We would reject the null hypothesis of no difference and say that the two low values 2.02 and 2.22 are not from the same population as the rest of this set of data. An extensive table of critical values for this test is included in Appendix A.

It would also be well to apply the appropriate Dixon ratio test to the inside value (either X_2 or X_{n-1}) of the two suspected outliers. If this observation is rejected, then the outside observation would, of course, be even more of an outlier.

III. REJECTION OF MULTIPLE OUTLIERS

The tests for the rejection of multiple outliers are theoretically more complex and therefore in practice more complicated than for a single outlier or a one-sided two-observation rejection test. Several of these tests have been developed in recent years.

BEST COPY AVAILABLE

TABLE 6.2. CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ or $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS*

Number of Observations n	10% Significance Level	5% Significance Level	1% Significance Level
4	.0031	.0008	.0000
5	.0376	.0183	.0035
6	.0921	.0565	.0186
7	.1479	.1020	.0440
8	.1994	.1478	.0750
9	.2454	.1909	.1082
10	.2853	.2305	.1415
11	.3226	.2666	.1736
12	.3552	.2996	.2044
13	.3843	.3295	.2333
14	.4106	.3568	.2605
15	.4345	.3818	.2859
16	.4562	.4048	.3098
17	.4761	.4259	.3321
18	.4944	.4455	.3530
19	.5113	.4636	.3725
20	.5269	.4804	.3909

*These significance levels are taken from Table V of Grubbs, Reference [8]. An observed ratio less than the appropriate critical ratio in this table calls for rejection of the null hypothesis.

This Table was taken from Grubbs 1969 page 11. By permission. Chapter 3, Army Statistics Manual, DARCOM p. 706-1-3.

Tietzen and Moore 1972[20] developed their test for two or more outliers where some are larger and some are smaller than the remaining values of the sample. This test would have difficulty in properly identifying the outliers of a set of 10 observations where the n−1 value is 10, and n value is 100 and $\bar{X}$ is 11.

The Rosner (1975)[18] test provides a solution for such a set of data. The test is set up to detect as many as k outliers. The value of k can be set wherever one desires. The value of k is used to trim the sample. That is, the k largest and k smallest values are omitted from the sample so that no outliers remain but only good values.

The test is then handled by calculating a "trimmed mean, a" and a "trimmed variance, b^2" as follows.

$$a = \sum_{i=k+1}^{n-k} \frac{Xi}{n-2k}$$

$$b^2 = \sum_{i=k+1}^{n-k} \frac{(Xi-a)^2}{n-2k-1}$$

Rosner uses these values to test the residual R, rather than the original $\bar{X}$ and s of the whole sample.

$$R_1 = \frac{\max\,|\,Xi-a\,|}{b}$$

This is compared to the critical value in Table 3. If the largest difference value is rejected then the second largest value is tested against the critical value in the same manner. The Rosner

BEST COPY AVAILABLE

TABLE 6.3A. PERCENTAGE POINTS OF ROSNER'S RST MANY OUTLIER TEST STATISTICS

R_1 AND R_2 k = 2

n	α =	0.10	0.05	0.01
10	R_1	7.35 ± .102	8.90 ± .146	13.38 ± .748
	R_2	4.92 ± .067	5.92 ± .103	9.13 ± .407
15	R_1	5.28 ± .063	6.01 ± .056	8.10 ± .208
	R_2	3.84 ± .045	4.31 ± .060	5.39 ± .134
20	R_1	4.64 ± .043	5.18 ± .053	6.47 ± .182
	R_2	3.50 ± .024	3.81 ± .032	4.70 ± .095
30	R_1	4.26 ± .027	4.62 ± .037	5.51 ± .108
	R_2	3.31 ± .021	3.57 ± .017	4.15 ± .053
40	R_1	4.04 ± .019	4.41 ± .033	5.26 ± .047
	R_2	3.23 ± .017	3.43 ± .030	3.92 ± .042
50	R_1	3.98 ± .013	4.25 ± .019	4.98 ± .081
	R_2	3.20 ± .011	3.39 ± .022	3.80 ± .047
75	R_1	3.89 ± .016	4.16 ± .016	4.77 ± .074
	R_2	3.19 ± .013	3.37 ± .029	3.72 ± .038
100	R_1	3.83 ± .016	4.09 ± .027	4.66 ± .088
	R_2	3.20 ± .012	3.34 ± .0076	3.74 ± .037

(The ± values are standard errors.)

This table is taken from Grubbs 1979, pages 3-65-67. By permission Grubbs table adapted from Rosner 1977.

BEST COPY AVAILABLE

TABLE 6.3B. PERCENTAGE POINTS OF ROSNER'S RST MANY OUTLIER TEST STATISTICS

R_1, R_2 AND R_3 k = 3

n	α =	0.10	0.05	0.01
20	R_1	5.91 ± .059	6.60 ± .079	8.19 ± .137
	R_2	4.50 ± .047	5.06 ± .052	6.34 ± .151
	R_3	3.73 ± .037	4.16 ± .046	5.22 ± .098
30	R_1	5.07 ± .037	5.60 ± .063	6.88 ± .093
	R_2	3.93 ± .028	4.32 ± .037	5.09 ± .121
	R_3	3.35 ± .016	3.62 ± .039	4.27 ± .076
40	R_1	4.60 ± .037	5.06 ± .040	6.05 ± .103
	R_2	3.68 ± .021	3.92 ± .021	4.53 ± .051
	R_3	3.20 ± .016	3.41 ± .024	3.82 ± .063
50	R_1	4.43 ± .033	4.76 ± .049	5.68 ± .038
	R_2	3.60 ± .014	3.82 ± .018	4.55 ± .086
	R_3	3.14 ± .019	3.30 ± .014	3.77 ± .047
75	R_1	4.18 ± .024	4.46 ± .034	5.10 ± .036
	R_2	3.47 ± .013	3.67 ± .019	4.10 ± .040
	R_3	3.08 ± .0096	3.19 ± .012	3.57 ± .045
100	R_1	4.12 ± .019	4.37 ± .034	4.98 ± .120
	R_2	3.44 ± .012	3.60 ± .022	3.88 ± .039
	R_3	3.10 ± .012	3.21 ± .016	3.45 ± .031

(The ± values are standard errors.)

BEST COPY AVAILABLE

TABLE 6.3C. PERCENTAGE POINTS OF ROSNER'S RST MANY OUTLIER TEST STATISTICS

R_1, R_2, R_3 AND R_4 $k = 4$

n	α=	0.10	0.05	0.01
20	R_1	7.56 ± .083	8.52 ± .112	11.70 ± .340
	R_2	5.88 ± .042	6.53 ± .050	8.83 ± .263
	R_3	4.91 ± .038	5.46 ± .064	7.23 ± .199
	R_4	4.17 ± .035	4.65 ± .056	6.03 ± .116
30	R_1	5.90 ± .030	6.40 ± .055	7.65 ± .096
	R_2	4.63 ± .030	5.01 ± .034	5.90 ± .094
	R_3	3.95 ± .037	4.27 ± .049	5.09 ± .089
	R_4	3.50 ± .024	3.76 ± .034	4.53 ± .101
40	R_1	5.23 ± .036	5.67 ± .066	6.85 ± .264
	R_2	4.13 ± .025	4.47 ± .037	5.24 ± .087
	R_3	3.60 ± .031	3.82 ± .030	4.52 ± .079
	R_4	3.25 ± .020	3.43 ± .027	3.99 ± .043
50	R_1	4.85 ± .036	5.19 ± .063	6.18 ± .111
	R_2	3.95 ± .022	4.18 ± .028	4.86 ± .082
	R_3	3.46 ± .014	3.67 ± .019	4.20 ± .066
	R_4	3.14 ± .0098	3.30 ± .021	3.75 ± .041
75	R_1	4.55 ± .039	4.87 ± .060	5.66 ± .105
	R_2	3.73 ± .022	3.94 ± .018	4.41 ± .054
	R_3	3.31 ± .010	3.47 ± .020	3.81 ± .021
	R_4	3.04 ± .014	3.16 ± .019	3.50 ± .034
100	R_1	4.43 ± .037	4.67 ± .034	5.38 ± .091
	R_2	3.64 ± .016	3.80 ± .018	4.28 ± .056
	R_3	3.27 ± .012	3.39 ± .011	3.72 ± .037
	R_4	3.03 ± .011	3.14 ± .012	3.41 ± .028

(The ± values are standard errors.)

BEST COPY AVAILABLE

sequential test is particularly powerful against outliers on one side rather than on both sides. It is flexible in that one does not have to specify an exact number of suspect outliers to be tested as was necessary in earlier tests.

An example application should clarify the approach. Use the Venus semi-diameter values cited earlier. The values in increasing order are: −1.40, −0.44, −0.30, −0.24, −0.22, −0.13, −0.05, 0.06, 0.10, 0.18, 0.20, 0.39, 0.48, 0.63 and 1.01.

1. Since the values −1.40 and 1.01 are suspect set k = 2

2. Discard the k lowest values: −1.40 and −0.44 and the k highest values; 0.63 and 1.01

3. Calculate the trimmed mean and standard deviation

$$a = \sum_{i=k+1}^{n-k} \frac{Xi}{n-2k} = 0.04273$$

$$b = \left(\sum_{i=k+1}^{n-k} \frac{(Xi-a)^2}{(n-2k-1)} \right)^{1/2} = 0.25$$

4. Calculate R1 and R2

$$R1 = \frac{\max |X_i - a|}{b} = \frac{-1.40 - 0.0427}{0.25} = 5.60$$

$$R2^* = \frac{\max |X_i - a|}{b} = \frac{+1.01 - .0427}{0.25} = 3.76$$

*Discarding X_i used in R_i value.

The critical values from Table 3 are:

For n = 15 α =	0.10	0.05	0.01
R1	5.28	6.01	8.10
R2	3.84	4.31	5.39

5. Conclude that the −1.40 value cannot be rejected at even the 5% level and the 1.01 value not even at the 10% level. These different conclusions seem to point out the weakness of the Rosner test when the outliers are on both sides.

Hawkins (1978)[15] has an effective multiple outlier test of the same type as the Grubbs test earlier. The test statistic E_k equals the inlier sums of squares, s^{2*}_K divided by the entire sample sums of squares s^2

$$E_k^* = \frac{s_k^{2*}}{s^2}$$

The critical values are given in Table 6.4.

For an example problem, we will use the Venus semi-diameter data again.

1. Calculate the inlier sums of squares (not including the − 1.40 and 1.01 values)

$$s_k^{2*} = .1936 + .09 + .0576 + .0484 + .0484 + .0169 + .0025 + .0036 + .01 + .0324 + .044 + .1521 + .2304 + .3969 = 1.27$$

$s_k^{2*} = 1.27$ for the n = 13 trimmed values

2. Calculate the sums of squares for the entire sample.

$$s^2 = 1.96 + .1936 + .09 + .0576 + .0484 + .0169 + .0025 + .0036 + .01 + .0324 + .04 + .1521 + .2304 + .3969 + 1.0201$$

$s^2 = 4.25$

3. Calculate the Hawkins test statistic

$$E_k^* = \frac{S_k^{2*}}{s^2} = \frac{1.27}{4.25} = 0.299$$

4. Conclude that both values should be rejected since the critical value from Table 6.4 for a k = 2, and an n of 15 at the 0.05 level is 0.3104. The critical value at the 0.01 level is 0.2319. The Hawkins test statistic E_k^* falls between these critical values.

BEST COPY AVAILABLE

TABLE 6.4. PERCENTAGE POINTS OF HAWKIN'S E_k^*

d.f.		0			10			20		
k	n/α =	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
2	5	0.0061	0.0011	0.0001	0.4556	0.3380	0.2205	0.6613	0.5610	0.4433
2	10	0.1640	0.0995	0.0487	0.4920	0.3985	0.2948	0.6461	0.5650	0.4664
2	15	0.3104	0.2319	0.1529	0.5404	0.4592	0.3638	0.6596	0.5890	0.5010
2	20	0.4136	0.3367	0.2509	0.5829	0.5105	0.4222	0.6780	0.6148	0.5345
2	25	0.4886	0.4168	0.3320	0.6186	0.5531	0.4712	0.6964	0.6388	0.5647
2	30	0.5455	0.4792	0.3982	0.6486	0.5887	0.5126	0.7135	0.6606	0.5917
2	40	0.6262	0.5698	0.4977	0.6959	0.6449	0.5782	0.7432	0.6975	0.6369
2	50	0.6810	0.6322	0.5684	0.7314	0.6869	0.6278	0.7676	0.7272	0.6731
2	75	0.7639	0.7277	0.6788	0.7907	0.7569	0.7111	0.8117	0.7804	0.7377
2	100	0.8109	0.7821	0.7428	0.8275	0.8002	0.7628	0.8412	0.8156	0.7803
3	10	0.0743	0.0395	0.0160	0.4000	0.3117	0.2190	0.5800	0.4959	0.3973
3	15	0.1967	0.1389	0.0847	0.4454	0.3684	0.2814	0.5860	0.5141	0.4270
3	20	0.2972	0.2338	0.1661	0.4892	0.4197	0.3376	0.6028	0.5386	0.4591
3	25	0.3758	0.3128	0.2410	0.5278	0.4641	0.3867	0.6217	0.5632	0.4896
3	30	0.4379	0.3776	0.3057	0.5612	0.5024	0.4292	0.6402	0.5863	0.5175
3	40	0.5297	0.4758	0.4084	0.6153	0.5642	0.4988	0.6737	0.6268	0.5657
3	50	0.5942	0.5463	0.4846	0.6570	0.6119	0.5530	0.7021	0.6604	0.6053
3	75	0.6948	0.6579	0.6087	0.7286	0.6936	0.6467	0.7551	0.7223	0.6782
3	100	0.7534	0.7236	0.6832	0.7741	0.7456	0.7067	0.7915	0.7645	0.7275
4	10	0.0299	0.0132	0.0041	0.3360	0.2515	0.1673	0.5357	0.4478	0.3485
4	15	0.1239	0.0819	0.0455	0.3753	0.3022	0.2225	0.5313	0.4583	0.3721
4	20	0.2154	0.1631	0.1099	0.4182	0.3516	0.2752	0.5449	0.4803	0.4019
4	25	0.2923	0.2370	0.1760	0.4575	0.3959	0.3227	0.5631	0.5043	0.4316
4	30	0.3558	0.3006	0.2368	0.4924	0.4350	0.3650	0.5819	0.5277	0.4595
4	40	0.4530	0.4015	0.3383	0.5504	0.4998	0.4360	0.6174	0.5700	0.5090
4	50	0.5235	0.4764	0.4169	0.5962	0.5509	0.4926	0.6483	0.6059	0.5506
4	75	0.6366	0.5992	0.5497	0.6765	0.6408	0.5933	0.7076	0.6739	0.6289
4	100	0.7041	0.6733	0.6319	0.7287	0.6992	0.6593	0.7492	0.7212	0.6832

This table is taken from Grubbs 1979 pages 3-69—71 by permission. Table adapted from Hawkins 1978. Chapter 3, Army Statistics Manual, DARCOM p. 706-1-3.

BEST COPY AVAILABLE

BEST COPY AVAILABLE

TABLE 6.4. PERCENTAGE POINTS OF HAWKIN'S E_k^* (Continued)

d.f.		0			10			20		
k	n/α =	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
5	10	0.0096	0.0033	0.0007	0.2906	0.2083	0.1308	0.5072	0.4139	0.3123
5	15	0.0763	0.0470	0.0236	0.3214	0.2518	0.1786	0.4891	0.4146	0.3290
5	20	0.1560	0.1138	0.0727	0.3618	0.2982	0.2271	0.4984	0.4333	0.3559
5	25	0.2283	0.1806	0.1296	0.4006	0.3413	0.2723	0.5151	0.4561	0.3843
5	30	0.2906	0.2412	0.1852	0.4359	0.3803	0.3136	0.5336	0.4792	0.4117
5	40	0.3895	0.3412	0.2828	0.4960	0.4463	0.3844	0.5699	0.5222	0.4615
5	50	0.4634	0.4182	0.3616	0.5443	0.4994	0.4421	0.6024	0.5596	0.5042
5	75	0.5854	0.5482	0.4994	0.6310	0.5950	0.5473	0.6662	0.6319	0.5863
5	100	0.6600	0.6289	0.5872	0.6885	0.6584	0.6180	0.7120	0.6832	0.6444
6	15	0.0451	0.0254	0.0112	0.2790	0.2121	0.1445	0.4562	0.3798	0.2943
6	20	0.1123	0.0784	0.0472	0.3160	0.2550	0.1887	0.4602	0.3945	0.3181
6	25	0.1786	0.1375	0.0949	0.3535	0.2964	0.2313	0.4748	0.4156	0.3448
6	30	0.2382	0.1938	0.1448	0.3886	0.3346	0.2710	0.4925	0.4381	0.3715
6	40	0.3364	0.2910	0.2371	0.4495	0.4007	0.3407	0.5289	0.4811	0.4209
6	50	0.4121	0.3685	0.3146	0.4995	0.4550	0.3988	0.5624	0.5193	0.4641
6	75	0.5403	0.5033	0.4551	0.5909	0.5546	0.5070	0.6296	0.5948	0.5488
6	100	0.6205	0.5890	0.5471	0.6526	0.6220	0.5811	0.6787	0.6493	0.6099
7	15	0.0249	0.0127	0.0049	0.2463	0.1821	0.1194	0.4381	0.3543	0.2685
7	20	0.0794	0.0534	0.0304	0.2786	0.2207	0.1592	0.4295	0.3636	0.2882
7	25	0.1389	0.1045	0.0700	0.3141	0.2597	0.1989	0.4410	0.3822	0.3128
7	30	0.1949	0.1562	0.1143	0.3481	0.2966	0.2367	0.4573	0.4033	0.3381
7	40	0.2907	0.2493	0.2007	0.4088	0.3617	0.3044	0.4929	0.4455	0.3864
7	50	0.3668	0.3261	0.2761	0.4596	0.4163	0.3620	0.5266	0.4839	0.4294
7	75	0.4992	0.4635	0.4172	0.5542	0.5184	0.4716	0.5960	0.5612	0.5155
7	100	0.5838	0.5530	0.5120	0.6192	0.5888	0.5481	0.6477	0.6182	0.5786

56

PERCENTAGE POINTS OF HAWKIN'S E_k^* (Continued)

d.f.		0			10			20		
k	n/α =	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
8	20	0.0546	0.0347	0.0182	0.2474	0.1913	0.1335	0.4047	0.3373	0.2618
8	25	0.1070	0.0780	0.0498	0.2805	0.2279	0.1703	0.4125	0.3530	0.2841
8	30	0.1591	0.1247	0.0884	0.3133	0.2632	0.2060	0.4270	0.3727	0.3081
8	40	0.2517	0.2130	0.1682	0.3732	0.3270	0.2715	0.4611	0.4136	0.3550
8	50	0.3275	0.2884	0.2409	0.4243	0.3815	0.3284	0.4947	0.4518	0.3977
8	75	0.4628	0.4273	0.3817	0.5212	0.4853	0.4387	0.5656	0.5305	0.4845
8	100	0.5510	0.5199	0.4789	0.5889	0.5581	0.5172	0.6193	0.6193	0.5494
9	20	0.0365	0.0217	0.0104	0.2210	0.1666	0.1123	0.3838	0.3146	0.2391
9	25	0.0818	0.0576	0.0351	0.2513	0.2005	0.1462	0.3874	0.3274	0.2591
9	30	0.1294	0.0992	0.0682	0.2827	0.2342	0.1798	0.3999	0.3455	0.2816
9	40	0.2178	0.1820	0.1413	0.3414	0.2963	0.2428	0.4324	0.3850	0.3270
9	50	0.2924	0.2553	0.2107	0.3924	0.3503	0.2986	0.4658	0.4228	0.3691
9	75	0.4290	0.3943	0.3499	0.4909	0.4551	0.4089	0.5377	0.5023	0.4564
9	100	0.5200	0.4891	0.4485	0.5608	0.5298	0.4888	0.5931	0.5627	0.5225
10	20	0.0230	0.0126	0.0054	0.1982	0.1455	0.0947	0.3671	0.2961	0.2204
10	25	0.0611	0.0414	0.0239	0.2258	0.1769	0.1258	0.3658	0.3055	0.2379
10	30	0.1043	0.0780	0.0518	0.2557	0.2089	0.1574	0.3761	0.3218	0.2589
10	40	0.1881	0.1551	0.1181	0.3130	0.2692	0.2178	0.4068	0.3596	0.3026
10	50	0.2612	0.2260	0.1842	0.3636	0.3225	0.2723	0.4396	0.3970	0.3439
10	75	0.3986	0.3645	0.3212	0.4632	0.4277	0.3822	0.5121	0.4768	0.4311
10	100	0.4920	0.4612	0.4209	0.5349	0.5039	0.4631	0.5688	0.5384	0.4982

BEST COPY AVAILABLE

IV. SKEWNESS AND KURTOSIS TESTS

These types of tests are routinely used to test for normality. The skewness test is used when suspect outliers are only on one side. The Kurtosis test is generally recommended when suspect outliers are on both sides or when scalar contamination is suspected. The tests can be applied sequentially until no more values are rejected. The articles by Ferguson 1961a[13] and Ferguson 1961b[14] and the critical value tables of Pearson 1965[17] or Mulholland 1977[16], should be consulted for more information on this approach.

V. SLIPPAGE TESTS

The foregoing tests have been based upon the assumption of dealing with basically a normal population. There are available distribution free or nonparametric tests which can be applied — Neave 1979.[9] The advantages of such tests are their protection against non-normally distributed populations. However, for a normal population, these tests are not as statistically powerful as are the parametric tests.

k — Sample Slippage Test — Conover 1971[1]

Consider the case where one has data from several, "k", random samples. These samples are mutually independent of each other as in the case of data generated at separate work-stations on a shop floor. Each sample also contains the same number of measurements, n.

The underlying assumptions are: 1) that each sample is a random sample from some population; 2) the k samples are mutually independent of each other; 3) the test will be exact for continuous random variables; 4) either the k population distribution functions are identical, or that some populations tend to yield larger observations than other populations (a one-sided test); and 5) the data are at least ordinal in scale.

The test operates by comparing the "largest" sample with the "smallest" sample in the following manner: Determine the largest measurement in each of the k samples and mark it with an asterisk. The smallest of these extreme values is designated z^1 and the largest z^k. The k-sample slippage test statistic Z equals the number of measurements from the sample containing z^k larger than the measurement z^1.

In the case where several of the k samples have the same z^k value, the second largest value should be used to break the tie. If necessary one could even go to the third level to designate the largest sample set z^k.

An example application foliows:

1. Ten measurements selected at random times from five separate (independent) work places are recorded below:

TABLE 6.5. WORKPLACE LOCATIONS

Measurements	A	B	C	D	E
1	143	210	159	113	98
2	152	183	69	128	106
3	216	258*	64	240*	66
4	236	204	85	134	177*
5	120	252	126	205	93
6	191	132	148	234	68
7	244*	171	98	188	81
8	104	231	90	149	143
9	208	226	202*	223	160
10	179	104	182	217	128

2. Place an asterisk beside the largest measurement in each sample (workplace location).

3. Underline the *smallest largest* measurement once (177 from location E).

BEST COPY AVAILABLE

4. Underline the *largest largest* measurement twice (258 from location B).

5. Count the number of measurements from location B that exceed the *smallest largest* value (177). This is seven. Therefore Z = 7.

6. From Table 6.6B the critical value for an $\alpha = .05 = 5$. Therefore, one concludes that workplace location B tends to have higher measurements than workplace E.

7. The test can be applied sequentially. In this case location A would be tested with a k of 4. It would also be determined to have slipped. Testing location D (k = 3) yields the decision that it too has slipped. Location C (k = 2) does not appear to have slipped.

This Conover k-sample slippage test has been studied for its power. It was thought to be more powerful than the parametric F test where the populations are normal and those populations with larger means also have larger variances.

TABLE 6.6A. QUANTILES OF THE *k*-SAMPLE SLIPPAGE TEST STATISTIC[a]

	k = 2					*k* = 3					*k* = 4				
	p = .80	.90	.95	.98	.99	*p* = .80	.90	.95	.98	.99	*p* = .80	.90	.95	.98	.99
n = 3	2	2				2									
4	2	3	3			3	3				3	3			
5	2	3	3	4	4	3	3	4	4		3	4	4		
6	2	3	4	4	5	3	4	4	5	5	3	4	4	5	5
7	2	3	4	5	5	3	4	4	5	5	3	4	5	5	6
8	2	3	4	5	5	3	4	4	5	6	3	4	5	5	6
9	3	3	4	5	5	3	4	5	5	6	3	4	5	6	6
10	3	3	4	5	6	3	4	5	6	6	4	4	5	6	6
12	3	3	4	5	6	3	4	5	6	6	4	4	5	6	7
14	3	3	4	5	6	3	4	5	6	7	4	5	5	6	7
16	3	4	4	5	6	3	4	5	6	7	4	5	5	6	7
18	3	4	4	5	6	3	4	5	6	7	4	5	6	7	7
20	3	4	4	6	6	3	4	5	6	7	4	5	6	7	7
25	3	4	5	6	6	3	4	5	6	7	4	5	6	7	8
30	3	4	5	6	7	3	4	5	7	7	4	5	6	7	8
35	3	4	5	6	7	3	4	5	7	7	4	5	6	7	8
40	3	4	5	6	7	3	4	5	7	8	4	5	6	7	8
Approximation for $n > 40$	3	4	5	6	7	4	5	6	7	8	4	5	6	8	9

[a]The entries in this table are p quantiles w_p, for selected values of p, of the slippage test statistic as defined in Section V. To obtain the p quantile, enter the portion of the table for the correct number of samples k, read across the row corresponding to the size n of each sample, to the entry w_p in the column headed with the desired value of $p = 1 - \alpha$. Reject H_0 at the level x if the test statistic exceeds w_p.

BEST COPY AVAILABLE

TABLE 6.6B. QUANTILES THE k-SAMPLE SLIPPAGE TEST STATISTIC (Continued)

	$k=5$					$k=6$					$k=7$				
	$p=.80$	.90	.95	.98	.99	$p=.80$	.90	.95	.98	.99	$p=.80$	.90	.95	.98	.99
$n=$ 4	3	3				3	3				3				
5	3	4	4			3	4	4			3	4	4		
6	3	4	4	5	5	4	4	5	5	5	4	4	5	5	
7	3	4	5	5	6	4	4	5	5	6	4	4	5	6	6
8	4	4	5	6	6	4	4	5	6	6	4	5	5	6	6
9	4	4	5	6	6	4	5	5	6	7	4	5	5	6	7
10	4	5	5	6	7	4	5	5	6	7	4	5	6	6	7
12	4	5	5	6	7	4	5	6	7	7	4	5	6	7	7
14	4	5	6	7	7	4	5	6	7	7	4	5	6	7	8
16	4	5	6	7	7	4	5	6	7	8	4	5	6	7	8
18	4	5	6	7	8	4	5	6	7	8	4	5	6	7	8
20	4	5	6	7	8	4	5	6	7	8	4	5	6	7	8
25	4	5	6	7	8	4	5	6	7	8	5	6	6	8	8
30	4	5	6	7	8	4	5	6	8	8	5	6	7	8	9
35	4	5	6	7	8	4	5	6	8	9	5	6	7	8	9
40	4	5	6	7	8	4	6	7	8	9	5	6	7	8	9
Approximation for $n>40$	4	6	7	8	9	5	6	7	8	9	5	6	7	9	10

TABLE 6.6C. QUANTILES OF THE k—SAMPLE SLIPPAGE TEST STATISTIC (Concluded)

	$k=8$					$k=9$					$k=10$				
	$p=.80$	.90	.95	.98	.99	$p=.80$	.90	.95	.98	.99	$p=.80$	.90	.95	.98	.99
$n=$ 4	3					3					3				
5	3	4	4			4	4	4			4	4	4		
6	4	4	5	5		4	4	5	5		4	4	5	5	
7	4	5	5	6	6	4	5	5	6	6	4	5	5	6	6
8	4	5	5	6	6	4	5	5	6	6	4	5	5	6	7
9	4	5	5	6	7	4	5	6	6	7	4	5	6	6	7
10	4	5	6	6	7	4	5	6	7	7	4	5	6	7	7
12	4	5	6	7	7	4	5	6	7	7	5	5	6	7	8
14	4	5	6	7	8	5	5	6	7	8	5	6	6	7	8
16	4	5	6	7	8	5	6	6	7	8	5	6	6	7	8
18	5	6	6	7	8	5	6	6	8	8	5	6	7	8	8
20	5	6	6	8	8	5	6	7	8	8	5	6	7	8	8
25	5	6	7	8	9	5	6	7	8	9	5	6	7	8	9
30	5	6	7	8	9	5	6	7	8	9	5	6	7	8	9
35	5	6	7	8	9	5	6	7	8	9	5	6	7	8	9
40	5	6	7	8	9	5	6	7	8	9	5	6	7	8	9
Approximation for $n>40$	5	6	7	9	10	5	6	8	9	10	5	7	8	9	10

These tables from k-Sample Slippage Test — Conover 1971. Practical Non-parametric Statistics, Conover ©1979. Reprinted by permission of John Wiley & Sons, Inc.

BEST COPY AVAILABLE

VI. YOUDEN EXTREME RANK SUM TEST FOR OUTLIERS

A non-parametric extreme rank sum test was originally published in 1963.[21] The mathematical development of this test was published by Thompson and Wilke also in 1963.[10] Data generated by multiple measuring instruments or people taking readings or measurements at several workplace areas would be a typical industrial hygiene application. Another application would involve evaluating the results of multiple readers on several asbestos filters to develop and qualify the personnel to be used for the procedure.

The Youden example adapted to the asbestos microscopic analysis problem is shown below.

The data point to Technician 2 as producing results lower than the other technicians.

This test is applicable to the case cited above, and of course, to the original work of Youden where temperature reference cells (objects) were tested (or judged) by several thermometers or where laboratories (objects) are tested by several samples (judges). Table 8 is presented from Youden and Steiner 1975.[11]

The application of the test to the above set of data follows.

1. Arrange the data as in Table 6.7.
2. Rank the values for each judge (thermometer, filter, material being tested) in ascending order. Ties can be broken by flipping a coin.
3. Sum up the ranks for each object being tested (chemical reference cell, technician, or laboratory).
4. Compare the ranks with the appropriate entry for Table 6.8. In this example the approximate 5% two-tail limits are 4 and 16 for the case of 4 objects and 4 judges.
5. With this small sample it would only be possible to reject the work of Technician 2. When he had the lowest or highest values on all four filters more filters would have to be examined.

TABLE 6.7. ASBESTOS MEASUREMENT VALUES

Technician	*Filter*				
	A	B	C	D	Rank Sum
1	36 (4)	38 (2)	36 (2)	30 (2)	10
2	17 (1)	18 (1)	26 (1)	17 (1)	4
3	30 (3)*	39 (3)	41 (4)	34 (4)	14
4	30 (2)*	45 (4)	38 (3)	33 (3)	12

*The tie was broken by flipping a coin.

BEST COPY AVAILABLE

TABLE 6.8. APPROXIMATE FIVE PERCENT TWO-TAIL LIMITS FOR RANKING SCORES

No. of Objects	Number of Judges												
No. of Labs.	*Number of Materials*												
	3	4	5	6	7	8	9	10	11	12	13	14	15
3		4	5	7	8	10	12	13	15	17	19	20	22
		12	15	17	20	22	24	27	29	31	33	36	38
4		4	6	8	10	12	14	16	18	20	22	24	26
		16	19	22	25	28	31	34	37	40	43	46	49
5		5	7	9	11	13	16	18	21	23	26	28	31
		19	23	27	31	35	38	42	45	49	52	56	59
6	3	5	7	10	12	15	18	21	23	26	29	32	35
	18	23	28	32	37	41	45	49	54	58	62	66	70
7	3	5	8	11	14	17	20	23	26	29	32	36	39
	21	27	32	37	42	47	52	57	62	67	72	76	81
8	3	6	9	12	15	18	22	25	29	32	36	39	43
	24	30	36	42	48	54	59	65	70	76	81	87	92
9	3	6	9	13	16	20	24	27	31	35	39	43	47
	27	34	41	47	54	60	66	73	79	85	91	97	103
10	4	7	10	14	17	21	26	30	34	38	43	47	51
	29	37	45	52	60	67	73	80	87	94	100	107	114
11	4	7	11	15	19	23	27	32	36	41	46	51	55
	32	41	49	57	65	73	81	88	96	103	110	117	125
12	4	7	11	15	20	24	29	34	39	44	49	54	59
	35	45	54	63	71	80	88	96	104	112	120	128	136
13	4	8	12	16	21	26	31	36	42	47	52	58	63
	38	48	58	68	77	86	95	104	112	121	130	138	147
14	4	8	12	17	22	27	33	38	44	50	56	61	67
	41	52	63	73	83	93	102	112	121	130	139	149	158
15	4	8	13	18	23	29	35	41	47	53	59	65	71
	44	56	67	78	89	99	109	119	129	139	149	159	169

This table is taken from the book, *Statistical Manual of the Association of Official Analytical Chemists* by W. J. Youden and E. H. Steiner 1975. By permission.

This test is a distribution free test as is the Conover-k-sample slippage test and hence should find considerable application where the industrial hygienist does not have a normal population(s) of measurements.

VII. SUMMARY

This outline discussed several techniques for dealing with two or more outliers in a sample. The handling of several sets of data was covered using the k-sample slippage test of Conover and the extreme rank sums test of Youden.

BEST COPY AVAILABLE

REFERENCES

1. Conover, W. J., *Practical Nonparametric Statistics*, John Wiley & Sons, Inc., New York, N.Y. 1971.
2. Grubbs, F. E., Sample Criteria for Testing Outlying Obssevations, Ann. Math., Statist. 21, 27—58, 1950.
3. Grubbs, F. E., Procedures for Detecting Outlying Observations in Samples, Technometrics 11, 1—21, 1969.
4. Grubbs, F. E.,Procedures for Detecting Outlying Observations, Chapter 3 of *Army Statistics Manual DARCOM-P706-1-3*, U.S. Army Aberdeen Research and Development Center, Aberdeen Proving Grounds, Maryland 21005, In Press 1979.
5. Grubbs, F. E., and Beck, G., Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations, Technometrics Vol. 14 No. 4, 847—854. 1972.
6. McMillan, R. G., Tests for One or Two Outliers in Normal Samples with Unknown Variance, Technometrics, Vol. 13, No. 1, 87—100, 1971.
7. McMillan, R. G., and David, H. A., Tests for One of Two Outliers in Normal Samples with Known Variance, Technometrics, Vol. 13, No. 1, 75—85, 1971.
8. Moran, M. A., and McMillan, R. G., Tests for One or Two Outliers in Normal Samples with Unknown Variance: A Correction, Technometrics, Vol. 15, No. 3, 637—40, 1973.
9. Neave, H. R., Quick and Simple Tests Based on Extreme Observations, Journal of Quality Technology, Vol. 11, No. 2, 66—79. 1979.
10. Thompson, W. A., and Willke, T. A., On an Extreme Rank Sum Test for Outliers, Biometrika 50, 375—83, 1963.
11. Youden, W. J., and Steiner, E. H., Statistical Manual of the Association of Official Analytical Chemists, AOAC, P.O. Box 540, Benjamin Franklin Station, Washington, D.C. 20044, 1975.

BEST COPY AVAILABLE

BIBLIOGRAPHY

12. Chauvenet, W., *A Manual of Spherical and Practical Astronomy:* Vol. II Dover Publications, Inc., New York, N.Y. (unabridged and unaltered republication of the fifth revised and corrected edition (copyright 1891)) 1960.

13. Ferguson, T. S., On the Rejection of Outliers, Fourth Berkeley Symposium on Mathematical Statistics and Probability, edited by Jerzy Neyman, University of California Press, Berkeley and Los Angeles. Pages 253–87, 1961.

14. Ferguson, T. S., Rules for Rejection of Outliers, Revue Institute Int. de Stat., Vol. 3, Pages 29–43, 1961.

15. Hawkins, D. M. Fractiles of an Extended Multiple Outlier Test, Manuscript, submitted for publication in the Journal of Statistical Computation and Simulation, 1978.

16. Mulholland, M. P., On the Null Distribution of $\sqrt{b_1}$ for Samples of Size at Most 25, with Tables, Biometrika, Vol. 64, No. 2, 401–409, 1977.

17. Pearson, E. S., Tables of Percentage Points of $\sqrt{b}$ and b_2 in Normal Samples; A Rounding Off, Biometrika Vol. 52, 282–285, 1965.

18. Rosner, B., On the Detection of Many Outliers, Technometrics, Vol. 17, No. 2, 221–7, 1975.

19. Rosner, B., Percentage Points for the RST Many Outlier Procedure, Technometrics Vol. 19, No. 3, 307–12, 1977.

20. Tietzen, G. L., and Moore, R. H., Some Grubbs-Type Statistics for the Detection of Several Outliers, Technometrics, Vol. 14, No. 3, 583–97, 1972.

21. Youden, W. J., Ranking Laboratories by Round Robin Tests, Material Research and Standards Vol. 3, 9–13, 1963.

BEST COPY AVAILABLE

APPENDIX A

TABLE OF CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ OR $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS

No. of Obs. n	Lower .1% Sig. Level	Lower .5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
4	.0000	.0000	.0000	.0002	.0008	.0031
5	.0003	.0018	.0035	.0090	.0183	.0376
6	.0039	.0116	.0186	.0349	.0564	.0920
7	.0135	.0308	.0440	.0708	.1020	.1479
8	.0290	.0563	.0750	.1101	.1478	.1994
9	.0489	.0851	.1082	.1492	.1909	.2454
10	.0714	.1150	.1414	.1864	.2305	.2863
11	.0953	.1448	.1736	.2213	.2667	.3227
12	.1198	.1738	.2043	.2537	.2996	.3552
13	.1441	.2016	.2333	.2836	.3295	.3845
14	.1680	.2280	.2605	.3112	.3568	.4106
15	.1912	.2530	.2859	.3367	.3818	.4345
16	.2136	.2767	.3098	.3603	.4048	.4562
17	.2350	.2990	.3321	.3822	.4259	.4761
18	.2556	.3200	.3530	.4025	.4455	.4944
19	.2752	.3398	.3725	.4214	.4636	.5113
20	.2939	.3585	.3909	.4391	.4804	.5270
21	.3118	.3761	.4082	.4556	.4961	.5415
22	.3288	.3927	.4245	.4711	.5107	.5550
23	.3450	.4085	.4398	.4857	.5244	.5677
24	.3605	.4234	.4543	.4994	.5373	.5795
25	.3752	.4376	.4680	.5123	.5495	.5906

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2; \ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i; \ x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$$

$$S^2_{1,2} = \sum_{i=2}^{n} (x_i - \bar{x}_{1,2})^2; \ \bar{x}_{1,2} = \frac{1}{n-2} \sum_{i=2}^{n} x_i$$

$$S^2_{n-1,n} = \sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2; \ \bar{x}_{n-1,n} = \frac{1}{n-2} \sum_{i=1}^{n-2} x_i$$

An observed ratio less than the appropriate critical ratio in this table calls for rejection of the null hypothesis.

The values are taken from Grubbs and Beck 1972. This table is taken from Grubbs 1979 pages 3-42—45. By permission. U.S. Army Statistics Manual. DARCOM, p. 706-1-3.

BEST COPY AVAILABLE

APPENDIX A (Continued)

TABLE OF CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ OR $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS

No. of Obs. n	Lower .1% Sig. Level	Lower .5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
26	.3893	.4510	.4810	.5245	.5609	.6011
27	.4027	.4638	.4933	.5360	.5717	.6110
28	.4156	.4759	.5050	.5470	.5819	.6203
29	.4279	.4875	.5162	.5574	.5916	.6292
30	.4397	.4985	.5268	.5672	.6008	.6375
31	.4510	.5091	.5369	.5766	.6095	.6455
32	.4618	.5192	.5465	.5856	.6178	.6530
33	.4722	.5288	.5557	.5941	.6257	.6602
34	.4821	.5381	.5646	.6023	.6333	.6671
35	.4917	.5469	.5730	.6101	.6405	.6737
36	.5009	.5554	.5811	.6175	.6474	.6800
37	.5098	.5636	.5889	.6247	.6541	.6860
38	.5184	.5714	.5963	.6316	.6604	.6917
39	.5266	.5789	.6035	.6382	.6665	.6972
40	.5345	.5862	.6104	.6445	.6724	.7025
41	.5422	.5932	.6170	.6506	.6780	.7076
42	.5496	.5999	.6234	.6565	.6834	.7125
43	.5568	.6064	.6296	.6621	.6886	.7172
44	.5637	.6127	.6355	.6676	.6936	.7218
45	.5704	.6188	.6412	.6728	.6985	.7261
46	.5768	.6246	.6468	.6779	.7032	.7304
47	.5831	.6303	.6521	.6828	.7077	.7345
48	.5892	.6358	.6573	.6876	.7120	.7384
49	.5951	.6411	.6623	.6921	.7163	.7422
50	.6008	.6462	.6672	.6966	.7203	.7459
51	.6063	.6512	.6719	.7009	.7243	.7495
52	.6117	.6560	.6765	.7051	.7281	.7529
53	.6169	.6607	.6809	.7091	.7319	.7563
54	.6220	.6653	.6852	.7130	.7355	.7595
55	.6269	.6697	.6894	.7168	.7390	.7627
56	.6317	.6740	.6934	.7205	.7424	.7658
57	.6364	.6782	.6974	.7241	.7456	.7687
58	.6410	.6823	.7012	.7276	.7489	.7716
59	.6454	.6862	.7049	.7310	.7520	.7744
60	.6497	.6901	.7086	.7343	.7550	.7772
61	.6539	.6938	.7121	.7375	.7580	.7798
62	.6580	.6975	.7155	.7406	.7608	.7824
63	.6620	.7010	.7189	.7437	.7636	.7850
64	.6658	.7045	.7221	.7467	.7664	.7874
65	.6696	.7079	.7253	.7496	.7690	.7898
66	.6733	.7112	.7284	.7524	.7716	.7921
67	.6770	.7144	.7314	.7551	.7741	.7944

BEST COPY AVAILABLE

APPENDIX A (Continued)

TABLE OF CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ OR $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS

No. of Obs. n	Lower .1% Sig. Level	Lower .5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
68	.6805	.7175	.7344	.7578	.7766	.7966
69	.6839	.7206	.7373	.7604	.7790	.7988
70	.6873	.7236	.7401	.7630	.7813	.8009
71	.6906	.7265	.7429	.7655	.7836	.8030
72	.6938	.7294	.7455	.7679	.7859	.8050
73	.6970	.7322	.7482	.7703	.7881	.8070
74	.7000	.7349	.7507	.7727	.7902	.8089
75	.7031	.7376	.7532	.7749	.7923	.8108
76	.7060	.7402	.7557	.7772	.7944	.8127
77	.7089	.7427	.7581	.7794	.7964	.8145
78	.7117	.7453	.7605	.7815	.7983	.8162
79	.7145	.7477	.7628	.7836	.8002	.8180
80	.7172	.7501	.7650	.7856	.8021	.8197
81	.7199	.7525	.7672	.7876	.8040	.8213
82	.7225	.7548	.7694	.7896	.8058	.8230
83	.7250	.7570	.7715	.7915	.8075	.8245
84	.7275	.7592	.7736	.7934	.8093	.8261
85	.7300	.7614	.7756	.7953	.8109	.8276
86	.7324	.7635	.7776	.7971	.8126	.8291
87	.7348	.7656	.7796	.7989	.8142	.8306
88	.7371	.7677	.7815	.8006	.8158	.8321
89	.7394	.7697	.7834	.8023	.8174	.8335
90	.7416	.7717	.7853	.8040	.8190	.8349
91	.7438	.7736	.7871	.8057	.8205	.8362
92	.7459	.7755	.7889	.8073	.8220	.8376
93	.7481	.7774	.7906	.8089	.8234	.8389
94	.7501	.7792	.7923	.8104	.8248	.8402
95	.7522	.7810	.7940	.8120	.8263	.8414
96	.7542	.7828	.7957	.8135	.8276	.8427
97	.7562	.7845	.7973	.8149	.8290	.8439
98	.7581	.7862	.7989	.8164	.8303	.8451
99	.7600	.7879	.8005	.8178	.8316	.8463
100	.7619	.7896	.8020	.8192	.8329	.8475
101	.7637	.7912	.8036	.8206	.8342	.8486
102	.7655	.7928	.8051	.8220	.8354	.8497
103	.7673	.7944	.8065	.8233	.8367	.8508
104	.7691	.7959	.8080	.8246	.8379	.8519
105	.7708	.7974	.8094	.8259	.8391	.8530
106	.7725	.7989	.8108	.8272	.8402	.8541

BEST COPY AVAILABLE

APPENDIX A (Continued)

TABLE OF CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ OR $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS

No. of Obs. n	Lower .1% Sig. Level	Lower .5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
107	.7742	.8004	.8122	.8284	.8414	.8551
108	.7758	.8018	.8136	.8297	.8425	.8563
109	.7774	.8033	.8149	.8309	.8436	.8571
110	.7790	.8047	.8162	.8321	.8447	.8581
111	.7806	.8061	.8175	.8333	.8458	.8591
112	.7821	.8074	.8188	.8344	.8469	.8600
113	.7837	.8088	.8200	.8356	.8479	.8610
114	.7852	.8101	.8213	.8367	.8489	.8619
115	.7866	.8114	.8225	.8378	.8500	.8628
116	.7881	.8127	.8237	.8389	.8510	.8637
117	.7895	.8139	.8249	.8400	.8519	.8646
118	.7909	.8152	.8261	.8410	.8529	.8655
119	.7923	.8164	.8272	.8421	.8539	.8664
120	.7937	.8176	.8284	.8431	.8548	.8672
121	.7951	.8188	.8295	.8441	.8557	.8681
122	.7964	.8200	.8306	.8451	.8567	.8689
123	.7977	.8211	.8317	.8461	.8576	.8697
124	.7990	.8223	.8327	.8471	.8585	.8705
125	.8003	.8234	.8338	.8480	.8593	.8715
126	.8016	.8245	.8348	.8490	.8602	.8721
127	.8028	.8256	.8359	.8499	.8611	.8729
128	.8041	.8267	.8369	.8508	.8619	.8737
129	.8053	.8278	.8379	.8517	.8627	.8744
130	.8065	.8288	.8389	.8526	.8636	.8752
131	.8077	.8299	.8398	.8535	.8644	.8759
132	.8088	.8309	.8408	.8544	.8652	.8766
133	.8100	.8319	.8418	.8553	.8660	.8773
134	.8111	.8329	.8427	.8561	.8668	.8780
135	.8122	.8339	.8436	.8570	.8675	.8787
136	.8134	.8349	.8445	.8578	.8683	.8794
137	.8145	.8358	.8454	.8586	.8690	.8801
138	.8155	.8368	.8463	.8594	.8698	.8808
139	.8166	.8377	.8472	.8602	.8705	.8814
140	.8176	.8387	.8481	.8610	.8712	.8821
141	.8187	.8396	.8489	.8618	.8720	.8827
142	.8197	.8405	.8498	.8625	.8727	.8834
143	.8207	.8414	.8506	.8633	.8734	.8840
144	.8218	.8423	.8515	.8641	.8741	.8846
145	.8227	.8431	.8523	.8648	.8747	.8853
146	.8237	.8440	.8531	.8655	.8754	.8859
147	.8247	.8449	.8539	.8663	.8761	.8865
148	.8256	.8457	.8547	.8670	.8767	.8871
149	.8266	.8465	.8555	.8677	.8774	.8877

BEST COPY AVAILABLE

Lesson 7

EXAMPLES AND EXERCISES – MULTIPLE OUTLIERS

TIME: Seventh Period, Second Day, (140 minutes)

OBJECTIVES: Enable the trainee to complete a set of examples and problems concerning data containing multiple outliers. This opportunity to practice will enable the trainee to exercise judgement as to the inclusion of the outliers regardless of their position in the data set.

EXHIBITS: Figure 7.1. Critical Value Diagram

Figure 7.2. Asbestos Plant Data

KEY TERMS: No new key terms in this lesson

MATERIALS: Student Manual

Pencil with Erasers

Calculators

Overhead Projector and Screen

Transparencies 7.1 and 7.2

METHOD: Practical Exercise

BEST COPY AVAILABLE

ABSTRACT

This lesson provides practice in using the concepts presented in Lesson 6. The following is an outline of the Examples and Exercises — Multiple Outliers Lesson:

- Introduction
- Exercises
- Solutions to Exercises

BEST COPY AVAILABLE

Lesson 7

EXAMPLES AND EXERCISES — MULTIPLE OUTLIERS

INTRODUCTION

In Lesson 6 you learned how to use a number of statistical tests together with their associated tables. These tests enable you to make a statistical decision regarding the treatment of two or more outliers.

The following problems will give you an opportunity to practice using these techniques.

1. Work by yourself unless the instructor directs otherwise.

2. Refer to Lesson 6 as the need arises.

3. Ask the instructor for help if you need it. Be prepared to discuss your solutions with the rest of the class.

EXERCISES

Complete the following problems following the directions for each one.

1. The data from a series of measurements were 4.4 ppm, 2.1, 5.0, 4.6, 5.0, 4.6, 4.9, 4.0, 10.6 and 4.3.

   Use the test statistic D_1 to determine if the value farthest from the mean should be rejected after using the D_3 test. If that value is rejected, the next largest difference should be tested again using test D_1.

2. Use the data set from a series of desorption efficiency measurements 81%, 83%, 94%, 93%, 95%, 97%, 96%.

   Determine whether the two lowest desorption efficiencies should be rejected.

3. Use the data set below from a desorption efficiency test, 86%, 88%, 93%, 94%, 95%, 96%, 97%.

   Determine whether the two lowest values should be rejected.

4. Use the data from problem 3 of Lesson 6. The observed values are 44 ppm, 21, 50, 46, 50, 46, 49, 40, 66, 43.

   Use the Rosner test to determine the presence of outliers.

5. Use the same set of data as for problem 4 and apply the Hawkins multiple outlier test to detect outliers.

6. Five operations in an asbestos plant were checked for their release of asbestos fibers to the general workplace environment. There were a total of ten samples taken at each operation randomly over the day for the same sampling time in all cases.

BEST COPY AVAILABLE

Asbestos Plant Work Location Measurements

Sample No.	A	B	C	D	E
1	3.6	4.2	5.1	13.7	4.5
2	8.7	9.9	6.2	9.9	3.9
3	2.1	10.7	7.5	11.0	8.7
4	0.7	8.9	4.1	9.6	4.5
5	5.3	9.1	2.1	5.7	8.8
6	4.7	4.5	10.1	8.6	3.5
7	6.9	9.4	11.5	10.7	7.0
8	4.3	3.7	2.9	12.7	9.0
9	2.1	2.1	8.0	9.8	5.5
10	5.3	11.0	6.0	8.0	6.4

Use the k-sample slippage tests to determine whether any locations are significantly different.

BEST COPY AVAILABLE

Lesson 8

CONTINUOUS DATA BEARING OUTLIERS

TIME: Third Period, Second Day (50 minutes)

OBJECTIVES: Enable the trainee to evaluate outliers from data generated continuously or continually by certain sampling methods.

EXHIBITS:

Table 8.1. Example Data Set

Figure 8.1. Order of Observations in Time

Table 8.2. Testing Randomness of Grouping in a Sequence of Alternatives (P = 0.005)

Figure 8.2. Testing Randomness in Grouping in a Sequence of Alternatives (P = 0.005)

Table 8.3. Testing Randomness in Grouping in a Sequence of Alternatives (P = 0.05)

Table 8.4. Limiting Values for the Total Number of Runs Above and Below the Median of a Set of Values

Table 8.5. Example Data Set

Table 8.6. Critical Values for the Lengths of Runs on Either Side of the Median of n Cases

Table 8.7. Limiting Values for Lengths of Runs Up and Down in a Series of n numbers

KEY TERMS:

Run up

Run down

Run above average

Run below average

Length of Run

BEST COPY AVAILABLE

MATERIALS: **Student Manual**

Chalk Board, Chalk and Eraser

Overhead Projector and Screen

Overhead Transparencies

Tables 8.1 through 8.7

Figure 8.1

METHOD: **Explanation/Lecture**

ABSTRACT

The following is an outline of the contents of the Continuous Data Bearing Outliers Lesson:

- Introduction
- Duplicate Measurements
- Routine Triplicate Measurements
- Control Charts
- Use of Runs
- Conclusion
- References

BEST COPY AVAILABLE

Lesson 8

CONTINUOUS DATA BEARING OUTLIERS

INTRODUCTION

The preceding discussions covered outlier rejection criteria and techniques from a single set of data or the rejection of a set of data from several sets of data. The techniques of Conover (k-sample slippage test) and Youden (extreme rank sum test) are examples of the latter. There are cases and processes that produce data either continually or continuously. These situations will be covered now.

I. DUPLICATE MEASUREMENTS

Replicate determination rejection rules were discussed by Anscomb 1960.[1] Tietjen and Beckman 1974[5] presented a comprehensive discussion of the analytical laboratory problem of dealing with duplicate measurements. If the two results do not agree, what happens? If a third analysis is run, how is its result handled? The same problem of two laboratories performing an assay or analysis arises. If the results disagree, and a referee laboratory runs an analysis, what should be done?

The various procedures that could be used are outlined below in case of disagreement between the two original results (either intralab or interlab).

A. Use the third analysis (or referee analysis) value only as the reported or accepted value.

B. Use the third analysis result (or referee value) if it falls within the first two. If not, accept the value closest to referee's result. This procedure is simply, accept the median value.

C. Use the average of the third value and the closest of the two original values.

D. Use the average of the closest pair of the three values.

E. Use the average of all three observations.

F. Use the average of the lowest pair. (This procedure is used where outliers, due primarily to contamination, can occur only on the high side).

G. Use the average of the highest and lowest values.

The decision to do a third analysis is in itself a rejection of outlier decision. Tietjen and Beckman recommend a Grubbs single sample test with some prior knowledge of the variability.

Tietjen and Beckman conclude that:

1. E and G are worst (heavily biased)
2. F is best if one can expect outliers to occur on only one side.
3. C and D are best and practically indistinguishable.
4. A and B are intermediate choices between the best (C and D) and the worst (E and G).

They conclude their studies with the statement "The final conclusion, restated because it is not in accord with common practice, is that with no prior knowledge of where the outliers

BEST COPY AVAILABLE

will occur, and given that the first two observations are not in agreement, it is uniformly better to report the mean of the closest pair than either the median or the referee's value."

II. ROUTINE TRIPLICATE MEASUREMENTS

Tietjen and Beckman summarize the arguments of Youden 1949[10] and Lieblein 1952.[6] The three measurements are a random sample from a normal population. The common practice of taking the mean of the "closest two" is condemned. The approved technique is to test for an outlier. If one is found at, say, the $\alpha \leq 0.01$ level, average the two remaining observations.

III. CONTROL CHARTS

The Shewhart control chart is a classical rejection of outlier technique. The $\bar{X}$ chart is specifically designed to detect values arising from a location change in the basic population being monitored. The range or R chart is specifically designed to detect a change in population variability or spread through the increased range of the sample values.

Since the "out-of-control" points of the control chart are by definition "outliers," no more need be said about this basic technique covered in earlier courses. Standard texts such as Grant 1964[4] and the ASTM Monograph on Control Charts[2] can be consulted.

IV. USE OF RUNS

A run is defined as a succession of values of the same type. When the values are given a two-way classification such as: go or no go; defective or non-defective; above-the-average or below-the-average; or increasing or decreasing.

The following set of values (adopted from Duncan 1959 page 118)[3] will illustrate the problem.

TABLE 8.1

Order of Sample	Value	Order of Sample	Value	Order of Sample	Value
In time		In time		In time	
1	13	13	8	25	10
2	10	14	2	26	11
3	14	15	5	27	16
4	9	16	8	28	11
5	6	17	9	29	9
6	9	18	10	30	4
7	16	19	17	31	9
8	12	20	12	32	20
9	5	21	14	33	18
10	9	22	27		
11	22	23	17		
12	12	24	9		

BEST COPY AVAILABLE

These values are plotted below on Figure 1.

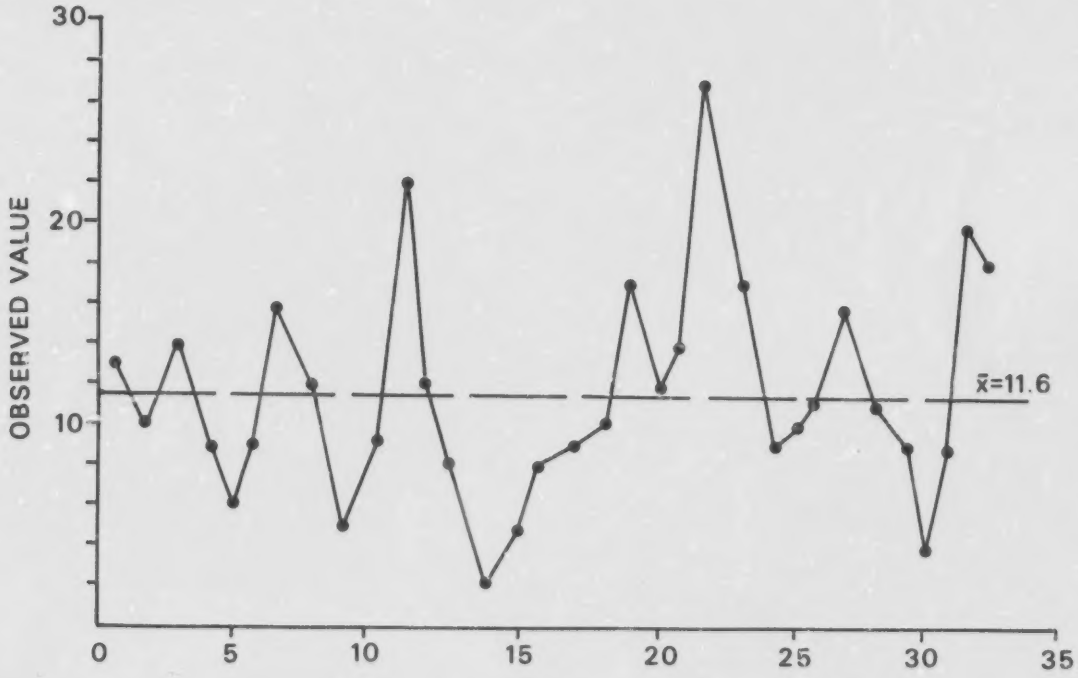


Figure 8.1

The types of runs that might be looked at in this set of data include "runs above average," "runs below average," "run up," "run down." Since the "run up" is a succession of increasing values, the positions 16, 17, 18, 19, are a "run up." Positions 12, 13, 14 are a "run down." Positions 19, 20, 21, 22, 23 are a "run above average" whereas positions 13, 14, 15, 16, 17, 18 are a "run below average."

Two approaches are taken in studying runs. The total number of runs of a type or the longest run of a type are both useful.

The total number of runs of a type from a random series have been studied by Swed and Eisenhart 1943,[8] Stevens 1939,[7] and Wald and Wolfowitz 1940.[9] The test is run as follows (as outlined by Duncan 1959).

1. Count the runs of one type, e.g., runs above the line:

Run Type	Positions	# of Runs
Runs of 1	(1),(3),(26)	3
Runs of 2	(7,8),(12,13),(32,33)	3
Runs of 5	(19,20,21,22,23)	1

2. Count the runs of the other type — runs below the line:

Run Type	Positions	# of Runs
Runs of 1	(2)	2
Runs of 2	(9,10)	1
Runs of 3	(4,5,6), (24,25,26)	2
Runs of 4	(28,29,30,31)	1
Runs of 6	(13,14,15,16,17,18)	1
		7

BEST COPY AVAILABLE

3. Let the smaller number of values of one type equal m. In this case the number of values above the line is 14 so m = 14.

4. Let the larger number of values (of the other type) equal n. In this example the number of values below the line is 19 so n = 19.

5. Determine the total number of runs of both types. In this case

   runs above the average = 7
   runs below the average = 7

   Total # of runs (observed) = 14

6. The values of m and n are used to enter the Tables 2 or 3 at either the P = 0.005 or P = 0.05 level to determine the critical values. If the observed total number of runs is less than the critical value we can conclude that the number of runs is less than one would expect based on the assumption of randomness. In this example the observed number of runs — 14 is not less than the critical values (9 at P = 0.005 or 12 at P = 0.05). Therefore we do not reject the hypothesis of randomness.

The same approach to examining the runs up and runs down yields that the number of increasing values equals 17; the number of decreasing values equals 15. The number of runs up equals 7 and the number of runs down equals 8. The total number of runs equals 15. Entering Tables 2 and 3 for m = 15, and

TABLE 8.2. TABLE FOR TESTING RANDOMNESS OF GROUPING IN A SEQUENCE OF ALTERNATIVES

(Probability of an equal or smaller number of runs than that listed is P = 0.005). n equals cases on one side of average — m equals cases on other side of average. m is always the smaller number of cases, n the larger.

n \ m	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
6	2														
7	2	3													
8	3	3	3												
9	3	3	3	4											
10	3	3	4	4	5										
11	3	4	4	5	5	5									
12	3	4	4	5	5	6	6								
13	3	4	5	5	5	6	6	7							
14	4	4	5	5	6	6	7	7	7						
15	4	4	5	6	6	7	7	7	8	8					
16	4	5	5	6	6	7	7	8	8	9	9				
17	4	5	5	6	7	7	8	8	8	9	9	10			
18	4	5	6	6	7	7	8	8	9	9	10	10	11		
19	4	5	6	6	7	8	8	9	9	10	10	10	11	11	
20	4	5	6	7	7	8	8	9	9	10	10	11	11	12	12

Table taken from Swed, F. S. & Eisenhart, C. "Tables for Testing Randomness of Grouping in a Sequence of Alternatives." Annals of Mathematical Statistics, Vol. XIV, 1943. By permission.

BEST COPY AVAILABLE

TABLE 8.3. TABLE FOR TESTING RANDOMNESS OF GROUPING IN A SEQUENCE OF ALTERNATIVES

(Probability of an Equal or smaller number of runs than that listed is P = 0.05) n = cases on one side of average; m = cases on the other side of average: m always taken as the smaller number of cases and n the larger.

n \ m	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
6	3														
7	4	4													
8	4	4	5												
9	4	5	5	6											
10	5	5	6	6	6										
11	5	5	6	6	7	7									
12	5	6	6	7	7	8	8								
13	5	6	6	7	8	8	9	9							
14	5	6	7	7	8	8	9	9	10						
15	6	6	7	8	8	9	9	10	10	11					
16	6	6	7	8	8	9	10	10	11	11	11				
17	6	7	7	8	9	9	10	10	11	11	12	12			
18	6	7	8	8	9	10	10	11	11	12	12	13	13		
19	6	7	8	8	9	10	10	11	12	12	13	13	14	14	
20	6	7	8	9	9	10	11	11	12	12	13	13	14	14	15

Swed & Eisenhart.

n = 17; The critical values are (9 at P = 0.005). Again, one cannot make a case for nonrandomness.

An extension of Tables 2 and 3 to larger numbers is provided in Table 4.

The second approach to studying runs is the "length of run" of a specific type. The technique involves the use of the median of the set of data. In case there are multiple values of the median these points are assigned to the side with the smaller number of values to balance the values on both sides. For an odd number of median values, one is ignored.

Having divided the data as above, determine the longest run on either side of the median and determine whether it is longer than the critical values of Table 6. If the critical values of Table 6 are exceeded, we conclude that nonrandom factors are affecting this population.

In the case of the earlier set of data the analysis would proceed as follows.

1. Review the data from Table 1 and determine the median and assign multiple values to the side with fewer values. Use Table 5 on page 81.

   The median value is 10 as there are 14 positions with values of 9 or less and 16 with values of 11 or more. The values in positions 2 and 25 were assigned to

BEST COPY AVAILABLE

the low side as the runs are short. The 10 value of position 18 is left unassigned.

2. Determine the longest run on either side. The values for positions (13, 14, 15, 16, 17) and (19, 20, 21, 22, 23) are 5 long runs.

3. Consult Table 6 to determine that the observed run lengths above exceeds the

TABLE 8.4. LIMITING VALUES FOR THE TOTAL NUMBER OF RUNS ABOVE AND BELOW THE MEDIAN OF A SET OF VALUES

Probability of an Equal or Smaller Value			Probability of an Equal or Smaller Value			Probability of an Equal or Smaller Value		
$m = n$	0.005	0.05	$m = n$	0.005	0.05	$m = n$	0.005	0.05
10	4	6	40	29	33	70	55	60
11	5	7	41	29	34	71	56	61
12	6	8	42	30	35	72	57	62
13	7	9	43	31	35	73	57	63
14	7	10	44	32	36	74	58	64
15	8	11	45	33	37	75	59	65
16	9	11	46	34	38	76	60	66
17	10	12	47	35	39	77	61	67
18	10	13	48	35	40	78	62	68
19	11	14	49	36	41	79	63	69
20	12	15	50	37	42	80	64	70
21	13	16	51	38	43	81	65	71
22	14	17	52	39	44	82	66	71
23	14	17	53	40	45	83	66	72
24	15	18	54	41	45	84	67	73
25	16	19	55	42	46	85	68	74
26	17	20	56	42	47	86	69	75
27	18	21	57	43	48	87	70	76
28	18	22	58	44	49	88	71	77
29	19	23	59	45	50	89	72	78
30	20	24	60	46	51	90	73	79
31	21	25	61	47	52	91	74	80
32	22	25	62	48	53	92	75	81
33	23	26	63	49	54	93	75	82
34	23	27	64	49	55	94	76	83
35	24	28	65	50	56	95	77	84
36	25	29	66	51	57	96	78	85
37	26	30	67	52	58	97	79	86
38	27	31	68	53	58	98	80	87
39	28	32	69	54	59	99	81	87
						100	82	88

Swed & Eisenhart.

BEST COPY AVAILABLE

critical values. The total number of points being considered is the n used in Table 6.

4. Since the critical value at the P = 0.05 is 8, no evidence of nonrandomness is present.

TABLE 8.5

Sample Order	Value	H or L Side
1	13	H
2	10	M-L
3	14	H
4	9	L
5	6	L
6	9	L
7	16	H
8	12	H
9	5	L
10	9	L
11	22	H
12	12	H
13	8	L
14	2	L
15	5	L
16	8	L
17	9	L
18	10	M-
19	17	H
20	12	H
21	14	H
22	27	H
23	17	H
24	9	L
25	10	M-L
26	11	H
27	16	H
28	11	H
29	9	L
30	4	L
31	9	L
32	20	H
33	18	H

TABLE 8.6. CRITICAL VALUES FOR LENGTHS OF RUNS ON EITHER SIDE OF THE MEDIAN OF *N* CASES

(Probability of getting at least one run of specified size or more)

N	0.05	0.01	0.001
10	5	--	--
20	7	8	9
30	8	9	--
40	9	10	12
50	10	11	--

F. Mosteller, "Note on Application of Runs to Quality Control Charts."Annals of Mathematical Statistics, Vol. XII, 1941, p. 232. By permission.

The final test for nonrandomness by runs involves studying runs up or down. The approach is similar to the preceding case.

1. Determine the length of the longest run up or down. The values for positions (15, 16, 17, 18, 19) represent the longest run, a run of 5.

2. Consult Table 7 to determine the probability of a given run length for the total number of values in the data set.

3. Since the probability of a run of 5 for an n of 33 appears to only approach the P = 0.05 level, one cannot conclude that the sample is nonrandom.

V. CONCLUSION

Rules are given for the proper consideration of duplicate and triplicate readings from an instrument or laboratory results. Proper handling of values after rejection is covered for these cases as well. Control charts are a natural outlier rejection technique specifically designed for the purpose in continuous data generating processes. The use of runs to determine nonrandomness provides an effective nonparametric technique.

BEST COPY AVAILABLE

TABLE 8.7. LIMITING VALUES FOR LENGTHS OF RUNS UP AND DOWN IN A SERIES OF *N* NUMBERS

Probability Equal to or Less Than 0.0032			Probability Equal to or Less Than 0.0567	
N	Run	Probability of an Equal or Greater Run	Run	Probability of an Equal or Greater Run
4	4	0.0028	4	0.0028
5	5	0.0004	4	0.0165
6	5	0.0028	4	0.0301
7	6	0.0004	4	0.0435
8	6	0.0007	4	0.0567
9	6	0.0011	5	0.0099
10	6	0.0014	5	0.0122
11	6	0.0018	5	0.0146
12	6	0.0021	5	0.0169
13	6	0.0025	5	0.0193
14	6	0.0028	5	0.0216
15	6	0.0032	5	0.0239
20	7	0.0006	5	0.0355
40	7	0.0015	6	0.0118
60	7	0.0023	6	0.0186
80	7	0.0032	6	0.0254
100	8	0.0005	6	0.0322
200	8	0.0010	7	0.0085
500	8	0.0024	7	0.0215
1,000	9	0.0005	7	0.0428
5,000	9	0.0025	8	0.0245

This table is taken from P. S. Olmstead, "Distribution of Sample Arrangements for Runs Up and Down." Annals of Mathematical Statistics, Vol. XVII (1946) p. 29. By permission.

REFERENCES

1. Anscombe, F. J., Rejection of Outliers, Technometrics Vol. 2, #2, pp. 123–147, 1960.

2. A.S.T.M. *Manual on Quality Control of Materials*, American Society for Testing and Materials, Philadelphia, Pa. 1951.

3. Duncan, A. J., *Quality Control and Industrial Statistics*, Richard D. Irwin, Homewood, Illinois, 1959.

4. Grant, E. L., *Statistical Quality Control*, 3rd ed., McGraw-Hill Book Co., New York, N.Y. 1964.

5. Tietjen, G. L. and Beckman, R. J. On Duplicate Measurements in the Chemical Laboratory, *Technometrics*, Vol. 16, #4, pp. 53–56, 1974.

BEST COPY AVAILABLE

BIBLIOGRAPHY

6. Lieblein, J., Properties of Certain Statistics Involving the Closest Pair in a Sample of Three Observations, *J. Research National Bureau of Standards, 48:*255—268, 1952.

7. Stevens, W. L., Distribution of Groups in a Sequence of Alternatives, Annals of Eugenics, Vol. IX, Part I, pp. 10—17, 1939.

8. Swed, S. and Eisenhart, C., Tables for Testing Randomness of Grouping in a Sequence of Alternatives, *Annals of Mathematical Statistics,* Vol. XIV, pp. 66—82, 1943.

9. Wald, A. and Wolfowitz, J., On a Test Whether Two Samples are from the Same Population, *Annals of Mathematical Statistics,* Vol. XI, pp. 147—162, 1940.

10. Youden, W. J., The Fallacy of the Best Two Out of Three.

BEST COPY AVAILABLE

BLANK PAGE

Lesson 9

EXAMPLES AND EXERCISES — CONTINUOUS DATA OUTLIERS

TIME: Ninth Period, Second Day (140 minutes)

OBJECTIVES: Enable the trainee to complete a set of examples and problems concerning continuous data which contain outliers. This opportunity to practice will give the trainee the ability to exercise judgement as to the inclusion of outliers using $\bar{X}$ - R charts or the analysis of the behavior of runs.

EXHIBITS: Figure 9.1. $\bar{X}$ — R Chart Worksheet

Figure 9.2. $\bar{X}$ - R Chart

KEY TERMS: No new key terms in this lesson.

MATERIALS: Student Manual

Pencils with Erasers

Calculators

Overhead Projector and Screen

Transparencies 9.1 and 9.2

METHODS: Practical Exercise

BEST COPY AVAILABLE

ABSTRACT

This lesson provides practice in using the concepts presented in Lesson 8. The following is an outline of the exercises contained in this lesson — Examples and Exercises, Continuous Data Outliers:

- Introduction
- Using an $\bar{X}$ - R Chart to Identify Outliers
- Analysis of Runs to Identify Outliers
- Solutions

BEST COPY AVAILABLE

Lesson 9

EXAMPLES AND EXERCISES – CONTINUOUS DATA OUTLIERS

INTRODUCTION

In Lesson 8 you learned how to use two techniques to identify outliers appearing in continuously generated data. The following problems will give you the opportunity to practice using these techniques.

1. Work by yourself unless the instructor directs otherwise.
2. Refer to Lesson 8 as the need arises.
3. Ask the instructor for help if you need it. Be prepared to discuss your solutions with the rest of the class.

EXERCISES

1. A certain manufacturer's industrial hygiene department calibrates its gas detector tube systems by checking two CO tubes from the same lot against a standard calibration gas concentration before each daily use. Over a period of a month the following results were recorded for concentration checks at the Federal standard.

   a. Enter the data on the Practice Worksheet, 9.1.

   b. Perform the calculations.

   c. Plot the data on the $\bar{X}$ (accuracy) and R (precision) charts, Worksheet 9.2.

Date	Sample A (CO in ppm)	Sample B (CO in ppm)
1	48	50
2	47	48
3	46	48
4	47	47
5	49	49
8	50	47
9	51	49
10	47	49
11	46	48
12	49	50
15	51	50
16	52	50
17	49	50
18	48	51
19	48	50
22	50	51
23	51	50
24	49	49
26	44	48
27	48	50

2. Examine the data from problem 1 above to determine the presence of runs and whether or not the data are randomly distributed.

BEST COPY AVAILABLE

Practice Worksheet – 9.1

Laboratory Quality Control $\bar{X}$–R Chart

Laboratory _____ Date _____

Method of Test or Operation _____

Reference Value _____ Increment of Measurement _____

_____

Data

Date	No.	X_1	X_2	X_3	$\bar{X}$	R
	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					
	11					
	12					
	13					
	14					
	15					
	16					
	17					
	18					
	19					
	20					

Totals $\Sigma \bar{X}$ _____ ΣR _____

X_i = observed value R = largest – smallest

k = sets of values CL = control limit

Σ = summation WL = warning limit

U = upper L = lower

D_4 = 3.267 for n' = 2; 2.575 for n' = 3

A_2 = 1.880 for n' = 2; 1.023 for n' = 3

n' = number of values in the set

Calculations

1. $\bar{\bar{R}} = \Sigma R \div k$

   ___ = ___ ÷ ___

2. $UCL_R = D_4 \times \bar{R}$

   ___ = ___ × ___

3. $UWL_R = 2/3(D_4\bar{R} - \bar{R}) + \bar{R}$

   ___ = 2/3(___ – ___) + ___

4. $\bar{\bar{X}} = \Sigma \bar{X} \div k$

   ___ = ___ ÷ ___

5. $CL_{\bar{X}} = A_2 \times \bar{R}$

   ___ = ___ × ___

6. $WL_{\bar{X}} = 2/3 \times CL_{\bar{X}}$

   ___ = 2/3 × ___

7. $UCL_{\bar{X}} = \bar{\bar{X}} + CL_{\bar{X}}$

   ___ = ___ + ___

8. $UWL_{\bar{X}} = \bar{\bar{X}} + WL_{\bar{X}}$

   ___ = ___ + ___

9. $LWL_{\bar{X}} = \bar{\bar{X}} - WL_{\bar{X}}$

   ___ = ___ – ___

10. $LCL_{\bar{X}} = \bar{\bar{X}} - CL_{\bar{X}}$

    ___ = ___ – ___

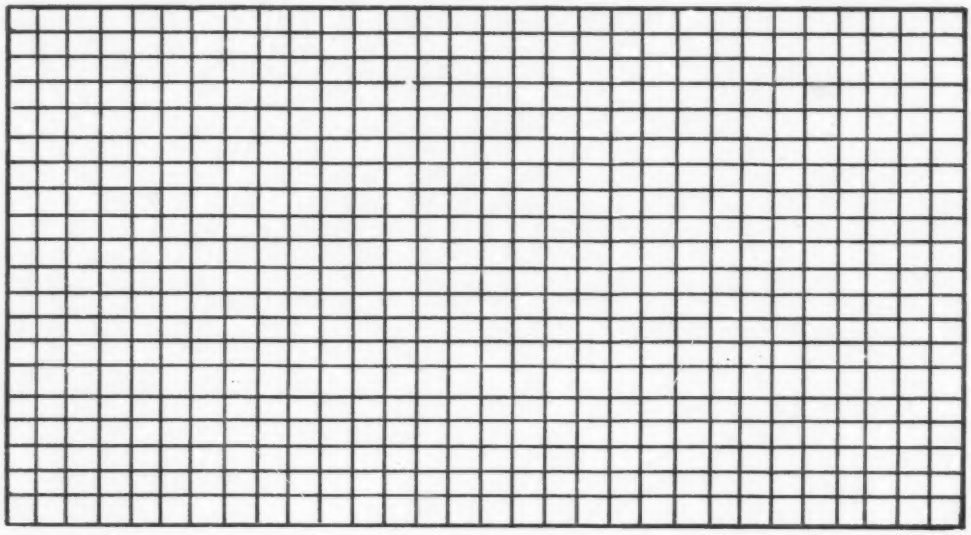
BEST COPY AVAILABLE

Practice Worksheet – 9.2

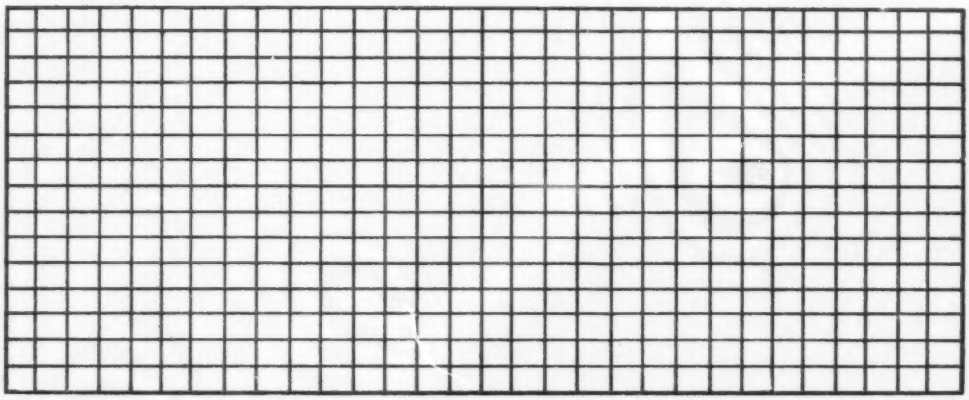
Laboratory Quality Control $\bar{X}$-R Chart

Operation _____ Date _____

Averages



Ranges



Sample Number

Directions:

Value	
1. Draw $\bar{R}$ line _____	6. Draw $UCL_{\bar{X}}$ line _____
2. Draw UCL_R line _____	7. Draw $UWL_{\bar{X}}$ line _____
3. Draw UWL_R line _____	8. Draw $LWL_{\bar{X}}$ line _____
4. Plot $\bar{R}$'s as generated	9. Draw $LCL_{\bar{X}}$ line _____
5. Draw $\bar{\bar{X}}$ line _____	10. Plot $\bar{X}$'s as generated

BEST COPY AVAILABLE

BLANK PAGE

Lesson 10

REVIEW AND EVALUATION

TIME: Tenth Period, Second Day (30 minutes)

OBJECTIVES: Enable the trainee to:

1. Provide feedback about the course
2. Emphasize topics of special importance or interest

EXHIBITS: None

KEY TERMS: No new key terms are introduced in this lesson.

MATERIALS: Student Manual

Pencils, Paper

METHODS: Discussion and Evaluation

BEST COPY AVAILABLE

ABSTRACT

The following is an outline of the contents of the Review and Evaluation Lesson:

- Introduction
- Course Evaluation
- Topic Emphasis

BEST COPY AVAILABLE

Lesson 10
REVIEW AND EVALUATION

INTRODUCTION

This is the final lesson of our course. There are two major tasks to accomplish, so we must move along quickly.

- **Course Evaluation**

  Here is your chance to say what you want about the course. A course evaluation form consisting of ten questions is included in this lesson for that purpose. Answer the questions on a separate sheet of paper or the NIOSH evaluation form; you do not have to identify yourself unless you wish. These sheets will be collected and sent to NIOSH in a sealed envelope. Neither your instructor nor your colleagues will see them.

- **Topics Emphasis**

  During the course, we may have missed or given little attention to a topic that is important to you. Be prepared to ask questions, share experiences, point out problems, or whatever. We will work in small groups in order to cover as much territory as possible.

Take 10 to 15 minutes to complete the Course Evaluation Form.

COURSE EVALUATION

Please answer the questions below on a separate sheet of paper. Number your answers.

1. What do you believe are the weaknesses of this course?
2. What do you believe are the strengths of this course?
3. Please comment on the instructors' presentations, describing what you liked and disliked about them.
4. What would you like to have added or expanded in the course?
5. What would you like to have eliminated or reduced in the course?
6. Did the course meet your expectations?
7. Please comment on the audio-visual aids describing what you liked and disliked about them.
8. Were the course arrangements (classroom facilities, lunchroom, accomodations, etc.) satisfactory?
9. Was the lesson time allotment suitable for the amount and difficulty of test material presented?
10. Any other comments, criticisms or suggestions?

TOPIC EMPHASIS

This portion of the Review and Review Lesson is devoted to an informal discussion session. If you haven't written down questions or noted problems during the course, take a few minutes now and jot some down. Next, break into groups according to the instructor's directions for the final discussion.

*USGPO: 1985-559-112-10949

BEST COPY AVAILABLE